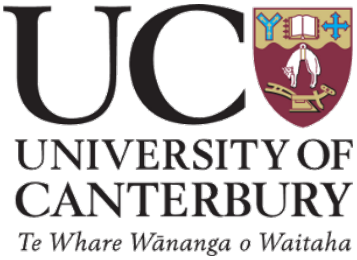
Evaluation of Deformable Image Registration



Joshua Bird

Department of Physics

University of Canterbury

Submitted to the University of Canterbury in partial fulfillment of the requirements for the degree of

Master of Science

2015

Acknowledgements

I would like to thank Isla Nixon, principle physicist at ADHB, for her immense support during the whole process, without which it would not have been possible to complete. I would also like to express my gratitude to my University supervisor Dr Steven Marsh for his helpful contributions without which I would be greatly disadvantaged. I would also like to thank my clinical supervisors Dr Richard Sims and Dr Gerard Bengua. Dr Richard Sims gave me a substantial amount of his time, advice and useful information during the development of the thesis and Dr Gerard Bengua gave me extensive and valuable guidance which I appreciate a great deal. I would like to thank my family and friends who gave me a lot of support and encouragement that helped me see it through to completion.

Abstract

Deformable image registration (DIR) is a type of registration that calculates a deformable vector field (DVF) between two image data sets and permits contour and dose propagation. However the calculation of a DVF is considered an ill-posed problem, as there is no exact solution to a deformation problem, therefore all DVFs calculated contain errors. As a result it is important to evaluate and assess the accuracy and limitations of any DIR algorithm intended for clinical use. The influence of image quality on the DIR algorithms performance was also evaluated.

The hybrid DIR algorithm in RayStation 4.0.1.4 was assessed using a number of evaluation methods and data. The evaluation methods were point of interest (POI) propagation, contour propagation and dose measurements. The data types used were phantom and patient data. A number of metrics were used for quantitative analysis and visual inspection was used for qualitative analysis.

The quantitative and qualitative results indicated that all DVFs calculated by the DIR algorithm contained errors which translated into errors in the propagated contours and propagated dose. The results showed that the errors were largest for small contour volumes (<20cm3) and for large anatomical volume changes between the image sets, which pushes the algorithms ability to deform, a significant decrease in accuracy was observed for anatomical volume changes of greater than 10%. When the propagated contours in the head and

neck were used for planning the errors in the DVF were found to cause under dosing to the target tumour by up to 32% and over dosing to the organs at risk (OAR) by up to 12% which is clinically significant. The results also indicated that the image quality does not significantly effect the DIR algorithms calculations. Dose measurements indicated errors in the DVF calculations that could potentially be clinically significant. The results indicate that contour propagation and dose propagation must be used with caution if clinical use is intended. For clinical use contour propagation requires evaluation of every propagated contour by an expert user and dose propagation requires thorough evaluation of the DVF.

Contents

List of Figures

List of Tables

List of Acronyms

ART Adaptive Radiation Therapy, 4

AVR Anatomical Volume Ratio, 13

CBCT Cone-Beam Computed Tomography, 5

CD Centroid Distance, 6

CTV Clinical Target Volume, 23

DIR Deformable Image Registration, 4

DSC Dice Similarity Coefficient, 6

DVF Deformable Vector Field, 5

DVH Dose Volume Histogram, 65

DVR Deformed Volume Ratio, 6

GTV Gross Target Volume, 23

ICRU International Commission on Radiation Units, 23

MU Monitor Unit, 31

OAR Organs at Risk, 2

POI Points of Interest, 6

PTV Planning Target Volume, 23

RO Radiation Oncologist, 23

ROI Regions of Interest, 7

RT Radiation Therapist, 23

TPS Treatment Planning System, 3

TRE Target Registration Error, 6

Chapter 1

Background and Introduction

1.1 Introduction

Cancer is a general term for a range of diseases, all of which are characterised by abnormal and malfunctioning cells within the body. Cells that have become cancerous are unable to produce tissue which functions normally within the body, instead, cancer cells grow and divide uncontrollably which leads to malignant neoplasms or tumours that may invade surrounding anatomy and metastasize to other sites in the body forming secondary tumours [37]. Cancer is one of the leading causes of death worldwide, accounting for 8.2 million deaths in 2012 which was approximately 15% of all deaths [13].

The standard treatments for cancer are surgery, chemotherapy or radiation therapy and are often done in conjunction with each other to improve treatment outcome. Radiation therapy is the use of ionising radiation in a controlled way to treat cancer patients, the treatments are either radical or palliative. Ionising radiation is used because of its ability to control the cell population of the targeted cancerous tumour, it does this by damaging the

genomic structures within the cell which induces cellular death [9].

The main aim of radiotherapy is to destroy the tumour tissue while minimising the damage to surrounding normal tissue which includes the organs at risk (OAR). The OARs are critical structures that have different radiosensitivities and if over dosed can cause significant toxicity effects to the patient. The therapeutic ratio is used to relate the tumour control probability to the normal tissue control probability, that is it relates the dose that is required to control or destroy the tumour tissue to the maximum dose the surrounding normal tissue can tolerate [5]. Normal tissues are at risk of getting damaged or destroyed by the doses required to control the tumour, however delivering insufficient dose to the tumour to protect the surrounding normal tissue limits the tumour control. The therapeutic ratio is used to maximise the control of the tumour tissue but at the same time minimise the damage to normal tissue. A range of ever advancing techniques and technologies are used to help target the tumour tissue better and therefore improve the therapeutic ratio [11] [4].

An important part of radiation therapy is medical imaging. Medical imaging is used to obtain anatomical and physiological information about a patient in order to extract relevant information for research or clinical purposes. This includes diagnosis, planning, treatment, guided treatment and disease monitoring [7]. The data provided by the different imaging modalities is a fundamental component to patient management in radiation therapy.

All interventions and associated procedures that require a patient to receive a dose of radiation must be justified and optimised according to the recommendations set by ICRP. The principle of justification is the requirement that any dose delivered to a patient does more good than harm. The principle of optimisation is the requirement that any dose delivered to a patient should be kept as low as reasonably achievable [24].

Radiation therapy treatments are generally delivered in fractions, this is based on the fact

that healthy cells tend to repair and regenerate faster than malignant cells so fractionation allows time for that to occur therefore minimising the detrimental effects on healthy cells. Tumour cells are also in different phases at different times, fractionation allows time for the different cells to move into a more radiosensitive phase [16].

As a result of fractionation the total treatment time is typically relatively long, with average time periods on the order of weeks during which time anatomical changes occur, the magnitudes of these changes are important and can influence the dose delivered and treatment outcomes [23]. The magnitudes of the anatomical changes can be assessed using medical imaging in conjunction with a treatment planning system (TPS).

In order to analyse medical images taken of the same patient, either by the same imaging modality (intra-modality) or by different imaging modalities (inter-modality), the images need to be aligned and fused to one another in a common coordinate system, this is commonly referred to as image registration. An image registration algorithm computes the required geometric transformations to map corresponding or homologous points between the two image sets to be co-registered [17]. In this thesis these are referred to as the reference image set and the target image set, registration results in the target image set being spatially aligned to the reference image set.

two images, a reference image and a target image, resulting in the reference image being spatially aligned to the target image. This process is generally done with images of the same scene from the same perspective, although not necessarily at the same time. The underlying goal of image registration is to accurately compare and enhance images that different imaging techniques can provide.

Image registration can be split into two broad categories, rigid and non-rigid (or deformable) registration. Rigid registration is a linear transformation that has 6 degrees of

freedom, 3 for translational (x, y, z), and 3 for rotational (pitch, yaw, roll). An extension to RR is the affine transformation which adds an extra 6 degrees of freedom for stretching/skewing in each dimension [12]. Deformable image registration (DIR) is a non-linear transform that increases the number of degrees of freedom and is often performed following a global RR. The accuracy and extent of the deformation depends on the transformation function used which range from a mesh grid that deforms according to higher order polynomials to a completely deformable (free-form or fluid) transformation. The free-form deformable registration generates a deformation vector field which carries information about the positional change of every voxel in the image. The vector fields point from the reference image voxel positions to the displaced voxel positions in the target image [21].

DIR has the potential to aid adaptive radiotherapy (ART). ART is where a radiotherapy treatment is adapted to suit the tissue changes of the patient, due to for example weight loss or swelling, as these tissue changes may have an adverse effect of the therapeutic ratio if they are not accounted for. ART is achieved by taking another computed tomography (CT) scan (re-CT) of the patient, in addition to the original CT scan. The re-CT image data is then transferred to a TPS for treatment replanning. DIR can potentially help aid this process by permitting contour propagation and dose propagation to the new CT image set which can aid in generating a new treatment plan [6].

Contour propagation is where the contours are deformed and transferred from the original planning CT images to the Re-CT images. The transfer of contours removes the need to re-contour (or at least can minimise the time spent re-contouring because any adjustments should be minimal, that is if the DIR transformations are accurate) and hence can potentially make the ART process more efficient [40].

Dose propagation is where dose is mapped from one CT image set to another CT image set using the deformable vector field (DVF) calculated by the DIR algorithm [29]. Dose

propagation can be used for dose accumulation, which is where the dose delivered from each treatment is summed or accumulated, which can potentially benefit ART. The idea behind dose accumulation is that the dose is calculated on each CBCT taken of the patient on treatment and mapped back to the original planning CT using the DVF calculated by the DIR algorithm, this is done for multiple CBCT scans and is then summed to get the dose accumulated to the patient over the course of treatment. This information could then potentially be used to make a decision on whether or not to pursue ART. This process requires that the DIR algorithm is able to calculate an accurate DVF, as the TPS uses this to map the dose from the CBCT to the planning CT, any error in the DVF will result in an error in the dose on the planning CT which undermines any decision to pursue ART.

Dose propagation is also important in the treatment of cancer recurrence whereby the dose is mapped from the previous treatment to the new one such that it is possible to reduce dose to critical OARs. Dose propagation is very sensitive to small deformation errors in high dose gradient regions and as such depends greatly on the deformation accuracy [18].

The calculation of a DVF is considered an ill-posed problem as there is no singular solution. Therefore if a DIR algorithm is being considered for clinical use it must be thoroughly assessed and evaluated. Multiple methods of assessment need to be undertaken, which include the use of clinical patient data, phantoms and dose measurements, this is required in order to establish the accuracy and limitations of the algorithm.

1.2 Aim

The main aim of this research is to assess and validate the hybrid DIR algorithm in the RayStation TPS for possible clinical use. The accuracy and limitations of the DIR algorithm

will be analysed. The effect of image quality on the DIR algorithm will also be evaluated along with the use of contour propagation and dose propagation which could potentially benefit ART.

The algorithm will be assessed using phantom CBCT and CT images and real patient CT images. The algorithm will be assessed by contour and POI propagation and by dose measurements. The contour and POI propagation will be analysed qualitatively using similarity metrics, these are the deformed volume ratio (DVR), the dice similarity coefficient (DSC), the centroid distance (CD) and the target registration error (TRE). The contour and POI propagation will also be analysed by visual inspection. The dose measurement data will be compared directly to the predicted dose on the TPS.

1.3 RayStation Algorithm

The DIR algorithm investigated was the hybrid DIR algorithm in RayStation version 4.0.1.4. The computation of a DVF requires a similarity measure which is the driving force of the computation. The similarity measure can be either intensity or geometrically based and in the case of the hybrid it can be both. The intensity based image similarity metric is the correlation coefficient. The geometric input is optional and can be a contour and/or points of interest (POI) defined on both image data sets, this helps guide the algorithm. The geometric inputs are called controlling structures and controlling POIs.

The DIR algorithm defines two image sets, the reference and the target. The deformation grid is defined as a set of points equally spaced on a lattice which divides space into box shaped elements or voxels. The DIR algorithm computes a vector field on the deformation grid, the vectors point from the individual voxels on the reference to the corresponding

voxels on the target image set. The vector field is defined as the deformation vector field (DVF).

The deformation vector at a grid point, xi, in three dimensional real-space is defined as, vi. The registration problem is formulated as a non-linear optimisation problem with the following objective function.

$$f(v) = \alpha C(v) + (\beta H(v) + \gamma S(v)\delta D(v)) \tag{1.1}$$

Where α, γ and δ are real non-negative weights. The term, C(v), is the intensity based image similarity function which uses the correlation coefficient to calculate a DVF which is modified and influenced by the other terms. The correlation coefficient measures the similarity between the deformed image set, which has had the DVF applied to it, and the target image set, this is used to try and increase the accuracy of the DVF calculated. The correlation coefficient is independent of linear transformations of the image intensities which makes it useful for the registration of CT and CBCT image sets [6] [39].

The term βH(v) + γS(v) controls the regularisation, which helps the algorithm calculate a realistic DVF. The first part of the regularisation term, βH(v), helps ensure that the DVF calculated is smooth such that there are no discontinuities in the vector field, this is done in part by essentially penalising large jumps in adjacent vector values [39]. The first part of the regularisation term also ensures that the DVF calculated is invertible, whereby the inverse DVF is also smooth and continuous, this is controlled by the parameter β, which is locally adjusted to achieve invertibility as illustrated in figure 1.1. The invertible requirement means that it is inverse consistent such that the algorithm gives a similar but inverted solution if the reference and target image sets were registered in the opposite direction [40]. The second part of the regularisation term, γS(v), uses the regions of interest (ROIs) as input to try and

aid the regularisation and increase the accuracy of the DVF.

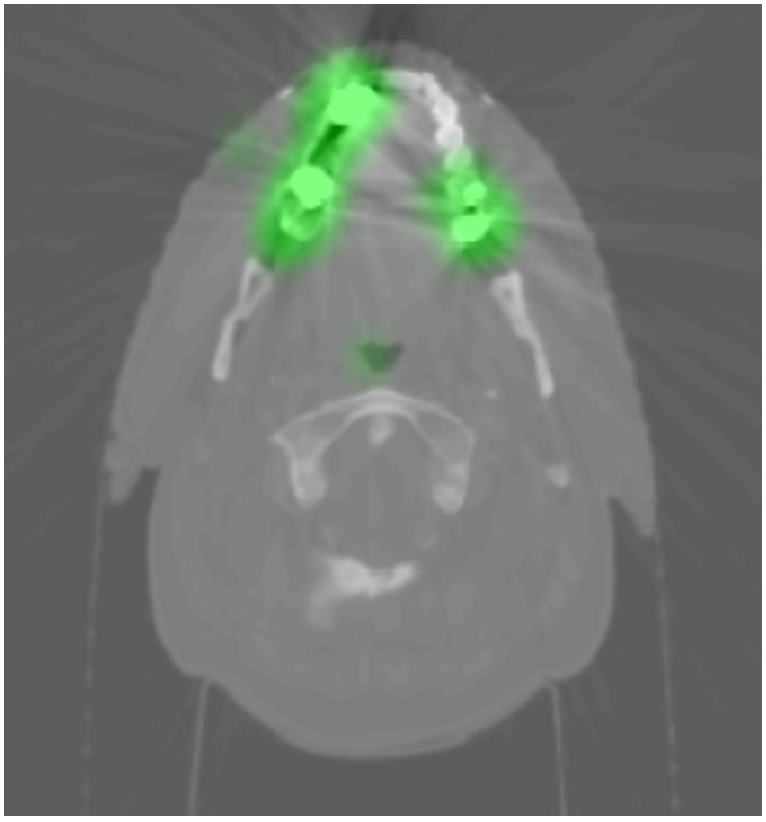


Fig. 1.1 Illustration of the adaptive regularisation weight for one CT-slice, the green highlights where the algorithm increased the regularisation weight to achieve local regularisation.

Any ROIs or POIs that have been input by the user are incorporated into the term, D(v), which is used to guide the algorithm [39]. A non-linear memory solver is used to compute the minimisation of the function, f(v), and solve for a DVF [8] [39]. The function, f(v), is optimised and solved using a resolution level system, whereby the DVF is calculated with a coarse resolution which is then used as the initial starting guess in the subsequent resolution level, this is completed iteratively until the finest level is reached [20] [33]. The algorithm has three resolution levels, an initial coarse resolution of 10mm, a subsequent resolution of 5mm and a final resolution of 2.5mm [39]. This approach to solving the function, f(v), has been shown experimentally to be more efficient than a single level approach [33].

1.4 Methods of Assessment

The DIR algorithm can be assessed in three main ways, these are propagation of ROI contours between registered images, propagation of identified points between registered images and measurement of dose propagation between registered images [6] [42]. Each evaluation method requires evaluation data, of which there are two types, phantom data and patient data [34]. Phantom data is obtained by scanning physical phantoms that can be deformed by a known amount. The phantom complexity can be relatively simple 2D block designs or more complex 3D anatomical phantoms. Phantoms used for this purpose typically attempt to represent a clinical scenario or at least an aspect of a clinical scenario such that the DIR performance can be related to the potential clinical use of the algorithm [22]. Phantom data is useful in that it provides a "gold standard" ground truth deformation which can be used to compare to the calculated deformation by the DIR algorithm. Patient data is the use of real patient CT images, which is useful as the DIR algorithm is intended to be used clinically on real patient CT image data.

Propagation of ROI contours between registered images is where regions of interest are delineated by an expert user on both the reference and target image set. An expert user is defined here as an individual that is able to delineate clinically acceptable contours. For phantom data the ROIs vary but are typically relatively easy to identify, however for patient data the ROIs are typically the tumour volume and OARs, however these structures vary depending on the anatomical site. The contours delineated on the reference image set, these are propagated to the target image, and are then compared to the expertly defined contours on the target image set. A range of metrics are used to compare the contour agreement and assess the accuracy of the propagated contours [6]. There are a couple advantages to using contour propagation as an evaluation method, one is that by increasing the number of contours in an image set you can increase the comprehensiveness of the evaluation, another

advantage is that contouring anatomical structures is a common practice in radiation therapy, therefore there are a number of expert users that can delineate clinically acceptable contours.

Propagation of identified points between registered images is where a number of POIs are identified on the reference image set and a number of corresponding POIs are identified on the target image set. The POIs are then propagated from the reference image set to the target image set and a number of metrics are used to compare and evaluate the agreement between the propagated POIs and the expertly defined POIs [6]. The ability to increase the information about the accuracy of the DVF by using more POIs is also an advantage for this evaluation method.

Measurement of dose propagation between registered images is where a plan is first created on the reference image set of a phantom arrangement, the dose is then calculated and deformed to the target image set phantom arrangement, finally the plan is delivered on the target image set phantom arrangement and the dose is measured and compared to the propagated dose [41]. The measurement of the dose can be made with a number of different types of detectors from point doses with an ion chamber through to 3D dosimetric deformable gel or "DEFGEL".

The methods described are intended to be different ways of evaluating the DVF calculated by the DIR algorithm, they also have the added benefit of directly evaluating the *contour propagation and dose propagation functions* of the DIR algorithm. The DIR algorithm will be assessed using all three methods described here to varying degrees and will utilise both phantom and patient data.

Chapter 2

Methods and Materials

2.1 Registration method

The rigid and deformable registration algorithms evaluated in this research were those used by RayStation. Described in this section is the detailed method used to register the image sets in RayStation.

Rigid Registration method

1. An external ROI which outlines the patient was defined on both the reference and target image set, this aids the rigid registration algorithm and is a pre-requisite for the deformable image registration algorithm. The external ROI was automatically defined using the “Create External ROI” tool.

2. Rigid image registration was performed between the reference and the target image

set with the automatic gray level based algorithm.

3. The algorithm has up to 6 degrees of freedom, 3 for translational (x, y, z) and 3 for rotational (pitch, yaw, roll). The rotational component can be set to zero if the registration only requires translational movement.

4. The registered images were qualitatively assessed by visual inspection.

DIR Registration method

1. A hybrid deformable image registration was performed between the reference and the target image set after the the rigid registration was complete using the method described previously. For the clinical studies the planning CT and re-CT image sets were selected as the reference and target image sets respectively. There were no controlling ROIs selected.

2. The contours, POIs and dose on the reference image set was propagated to the target image set for the appropriate studies using the *map ROIs, POIS and dose functions* for quantitative and qualitative assessment using the methods and metrics described in the following sections.

2.2 Metrics

2.2.1 Deformed Volume Ratio (DVR)

The deformed volume ratio (DVR) is defined in the following equation:

$$DVR = \frac{V_{expert}}{V_{propagated}} \tag{2.1}$$

Where V_{expert} is the volume defined by an expert on the target image set, and is considered the gold standard, and $V_{propagated}$ is the volume propagated by the DIR algorithm from the reference image set to the target image set [10].

Ideally this ratio would be unity, indicating that the expertly defined contour and the propagated contour are the same volume. The DVR indicates differences in size between the expertly defined contour and the propagated contour, however it does not measure the quality of the contours as two contours could have an identical volume but have a different shape and position. This metric is useful however, because it gives additional information about the similarity between the expertly defined contour and the propagated contour.

2.2.2 Anatomical Volume Ratio (AVR)

The anatomical volume ratio (AVR) is defined by the following equation:

$$AVR = \frac{V_{ER}}{V_{ET}} \tag{2.2}$$

Where V_{ER} is the volume defined by an expert on the reference image set, and V_{ET} is the volume defined by an expert on the target image set. The expertly defined contours are considered the gold standard and can be considered the true volume of the anatomical structure. This metric provides the true relative anatomical change from the reference CT image set to the target CT image set. The AVR metric is useful when assessing the impact of true relative anatomical volume changes on the algorithms ability to deform.

2.2.3 Dice Similarity Coefficient (DSC)

The DSC is a metric used to quantify the similarity between two volumes, in this case the DSC was used for quantifying the similarity between contoured regions of interest (ROI) on the target CT image data set.

The DSC is defined by the following equation:

$$DSC = \frac{2 \times A' \cap B}{A' + B} \tag{2.3}$$

Where A' is the deformed contour from the reference image set to the target image set, and B is the contour expertly defined on the target image set.

If there is no overlap between the two contours the DSC value is 0, if there is partial overlap the DSC is between 0 and 1, and if there is complete overlap between the volumes the DSC value is 1 [43]. The DSC is therefore restricted to the range [0, 1]. The values and corresponding contour spatial volume overlap is illustrated in Figure 2.1

The DSC is sensitive to both size and location of the contours. For example, two contours with the same volume but with only half the volume overlapping would result in a

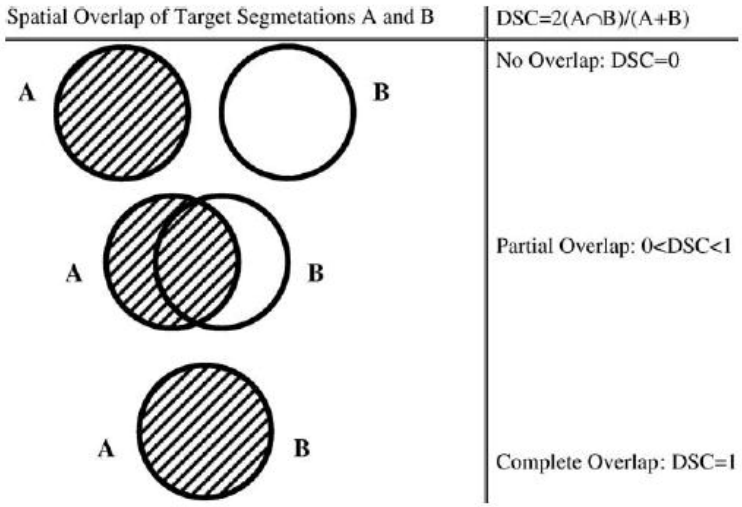


Fig. 2.1 Illustration of DSC [43].

DSC = $\frac{1}{2}$ and two contours where one is half the volume of the other but completely encompassed by the larger contour would result in a DSC = $\frac{2}{3}$. From this example it is clear that differences in location is more heavily weighted than differences in size.

The value of the DSC can be used to compare contour agreement across different image sets such that the DSC computed for the same ROI across multiple patients can be compared. The value of the DSC is an indicator of contour agreement and the ability to compare across multiple image sets makes the DSC a useful tool in deformable image registration assessment.

2.2.4 Centroid Distance (CD)

The centroid is the geometric center of a contour. The centroid distance metric is simply the distance from the centroid of the propagated contour to the centroid of the expertly defined contour on the same image set.

The centroid position is given in x, y, z coordinates, the distance to another centroid is

calculated using the following equation which is derived from the Pythagorean Theorem:

$$CD = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{2.4}$$

The centroid distance metric is used as an indicator of the quality and accuracy of the registration algorithm being assessed. The quality and accuracy is indicated by how close the centroid distance metric is to zero. Ideally the centroid distance metric would be zero, this would indicate that the contours are identical and in the same position. However it is possible to get a value of zero with contours that do not match, this happens when the centroids overlap coincidentally. It is however still a useful metric for providing information about the contours and is used as an additional metric to analyse contour similarity.

2.2.5 Target Registration Error (TRE)

The target registration error (TRE) is defined by the following equation:

$$TRE = \frac{1}{n} \sum_{n}^{i=1} T(X_i) - Y_i \tag{2.5}$$

Where T(X_i) is the point of interest (POI) on the reference image set that has had the DIR transformation applied to it, and Y_i is the corresponding POI on the target image set. The difference between T(X_i) and Y_i is summed for n POI and divided through by n to get the average. X_i and Y_i are vectors, the TRE is therefore a vector [6].

Corresponding POI are identified on both the reference and target image set which are in identical locations relative to the subject of the scan. The points on the reference image

set are propagated to the target, the euclidian distance between the deformed and the target points is then calculated and averaged to give the TRE.

2.3 Phantom Study

2.3.1 Multiple layer solid water phantom: without user input

A relatively simple phantom arrangement was used to simulate a volume change. Four layers of solid water were used to construct the phantom. Solid water was selected because it is a tissue equivalent material with absorbing and attenuation properties which are similar to soft tissue, having a density of 1 g/cm^3 with a corresponding CT number of approximately 1000 HU. The properties of solid water make it ideal for tissue change simulation.

The individual layers were all 30 cm by 30 cm in length and width with a varying thickness of 2, 1, 0.5 and 0.3 cm. Each layer was stacked uniformly on top of each other with the edges properly aligned. A scan was performed with and without the 0.3 cm layer which gave a scan with four layers and a scan with three layers. The volume with four layers was 3420 cm^3 and the volume with three layers was 3150 cm^3. Figure 2.2 shows the arrangement and the relative thickness of the different layers.

This phantom arrangement enabled the testing of the accuracy and consistency of the DIR algorithm. The DIR algorithm was used to register the CT image sets in both directions according to the registration method in section 2.1. Using this phantom arrangement there was a change in volume of 270 cm^3 between the two scans and within that volume the average CT number changes from $\sim$1000HU to $\sim$0 HU. This simulated both an expansion and contraction of tissue volume which could represent swelling or tumour reduction for

Fig. 2.2 Solid water phantom layers arrangement

example. In total there were 8 DVFs calculated, 4 contraction simulations and 4 expansion simulations.

Quantitative evaluation was made using the DSC and POI propagation. The DSC of the rigid registration was calculated as well as the average DSC for the contraction and the expansion simulations. A paired, two-tailed t-test was used to see if there was any statistically significant difference between the expansion and contraction DSC. All t-tests are assuming a 95% confidence level. If the algorithm is inverse consistent then the values of the contraction and expansion DSC should be approximately the same. Qualitative assessment was also made by visually assessing the DVF overlaid with the CT images to determine where the algorithm was correctly and incorrectly identifying changes and was also correlated with the quantitative results.

2.3.2 Multiple layer solid water phantom: with user input

The volume of CT voxels that the algorithm includes for its calculation of the DVF is larger than the volume of the phantom and includes part of the CT couch as well as air. This means that the CT couch influences the algorithm's calculations which can cause errors in the DVF. The automatically defined external contour was used as a controlling structure. A better DVF should be calculated by doing this, as the algorithm will concentrate on the phantom volume and will apply a lower weighting to the surrounding voxels in its calculation of the DVF.

2.3.3 Effect of Image quality on DIR

CT and CBCT scans rely on x-ray photons which is ionising and therefore delivers an absorbed dose to the patient [19]. The image quality directly depends on the number of x-ray photons and therefore all CT and CBCT scans are a trade-off between the image quality and dose [14]. The dose that the patient receives is determined by the scan parameters and the size of the patient. The scan parameters are set for the site and size of the patient such that they produce clinically acceptable image quality and dose to the patient [7]. However there are still image quality variations between patients for a given site and dose, also there are significant image quality differences between CT and CBCT, with CBCT having poorer image quality [31]. Therefore it is important that the influence of image quality is assessed when investigating deformable image registration.

It is expected that image quality will affect the DVF calculation, however the magnitude of the influence is unknown. Image quality has a number of components, in this study image noise and contrast were changed and their effect on the DVF was investigated. The level

of noise and contrast was controlled on CT by changing the kV and mAs parameters. The image quality was characterised on CT and CBCT by scanning the CATPHAN phantom.

The protocols in the CATPHAN manual were used to analyse the scans for image noise, contrast and resolution [15]. The image noise is indicated by the standard deviation of the mean Hounsfield unit (HU) measured on the image uniformity module of the CATPHAN phantom. The contrast is indicated by the ability to discern spheres of varying sizes and of varying grayscale on the contrast resolution module. The resolution is determined by the number of line pairs that can be resolved from a set of line pairs of varying space between each line on the resolution module [15].

The CT scans were done on the Siemens Sensation Open CT scanner. The CT scans were made such that the lowest, normal and highest possible image quality settings were selected. For the lowest, normal and highest image quality settings the kV was 80, 140 and 160 respectively and the mAs were 49, 160 and 357 respectively. The scan on the CBCT was done using the pelvis scan and had a kV and mAs of 125 and 680 respectively [35].

Solid water phantom: image quality assessment

The influence of image quality on the deformable image registration algorithm was assessed by using four layers of solid water. The individual layers were all 30 cm by 30 cm in length and width with a varying thickness of 2, 1, 0.5 and 0.3 cm. Each layer was stacked uniformly on top of each other with the edges properly aligned. The 0.3 cm layer was removed, resulting in a three layer phantom, which was then scanned again. Three different image quality settings on CT were used which resulted in three paired CT scans.

A DVF was calculated for each of the three paired CT scans of different image quali-

ties. The images were registered in both directions which gave a simulated contraction and expansion and resulted in 6 DVFs. For the simulated expansion, the reference was the three layer phantom and the target was the four layer phantom. The simulated contraction had the reverse arrangement with the reference being the four layer phantom and the target being the three layer phantom.

The deformations were assessed quantitatively using the DSC metric as defined in section 2.2.2 and qualitatively by visual inspection of the deformed images.

Apple phantom: image quality assessment

The influence of image quality on the deformable image registration algorithm was assessed by using an apple that had an applied deformation. An apple was used because it has CT numbers that fall within the range of soft tissue ($\sim$1000 HU) and it's CT numbers are non-uniform due to the non-uniform density, this gave a more realistic phantom for assessing the DIR algorithm. The apple was setup in an identical position on both CT and CBCT and was scanned in an un-deformed and deformed state. The CT scans were made with three different image qualities, lowest, normal and highest.

The apple had a known deformation applied, whereby part of the apple was removed. The known deformation was characterised by the external contour and POI. The apple was scanned in the deformed and un-deformed state for the three CT image qualities and the CBCT. This resulted in eight scans, six for CT and two for the CBCT. Each CT scan was registered to the appropriate CBCT scan. This resulted in six DVFs, three for the un-deformed CT to the deformed CBCT, and three for the deformed CT to the un-deformed CBCT.

This arrangement enabled the simulation of the expansion and contraction of soft tissue

which commonly occurs during treatment. For the simulated contraction the CT was the un-deformed apple and the CBCT was the deformed apple. For the simulated expansion the CT was the deformed apple and the CBCT was the un-deformed apple.

Deformable registration of CT to CBCT is of clinical interest as this will be necessary for dose accumulation which is used to assess the need for adaptive radiation therapy. It is therefore important to assess the influence of image quality on DIR and to ensure that the deformations are accurate.

The deformations were assessed quantitatively using the DSC metric as defined in section 2.2.2, the TRE metric as defined in section 2.2.4 and qualitatively by visual inspection of the deformed images.

2.4 Patient Study

A number of sites were selected for the general patient image set study, these were, the lung, breast, head and neck, cranialspinal, prostate and groin. The sites were selected based on the potential application of DIR to those sites with the use of advancing radiation therapy techniques.

The head and neck site was then selected as a further study because there is increasing and particular interest in using advanced techniques involving DIR for this treatment site.

2.4.1 General patient image sets: without user input

For this retrospective study, a cohort of 8 patients were selected from the Pinnacle TPS database and imported into RayStation. Each patient was selected based on the following criteria, first they had to have two image data sets, the original planning CT scan (CT1) and a re-CT scan (CT2). Secondly each patient had to have contours that matched on both of their data sets.

All contours were delineated at the time of CT1 and CT2 scans in order to facilitate treatment planning and assess the need to re-plan respectively. Once the initial planning CT scan was completed, each patient's CT image data were sent to the Pinnacle TPS for external beam planning. Standard contours were delineated as defined in the International Commission on Radiation Units (ICRU) Reports No. 50 and 62. These included the gross tumour volume (GTV), the planning target volume (PTV), the clinical target volume (CTV) and the OARs. The contours delineated on the second CT scan were variable, and did not conform exactly to the ICRU reports; this was because they only included clinically relevant structures for assessing the change in dose distribution from which a radiation oncologist (RO) made a decision on whether to re-plan or not. However there was still a reasonable range of contours that could be used for the study.

The patients selected had a range of treatment sites, and therefore a range of different contours, the same contours were not delineated on every patient, however each contour studied was delineated on a minimum of two patients for comparison. The contours included in this study were the PTV, lung, spinal cord, brainstem, mandible, bladder and rectum. For all treatment sites the PTV was contoured by a RO and the OAR were contoured by a radiation therapist (RT)[25][26][28][27].

The contours were propagated in 3 different ways, first using rigid registration, then DIR

and finally DIR with no rotation (DIR-NR). DIR with no rotational degrees of freedom was an additional registration method and was added due to a known error in the DIR algorithms inability to take into account the rotational component of the rigid registration when computing contour or dose propagation [30]. All contours were propagated according to the rigid registration and DIR method in section 2.1 and assessed quantitatively using the volume difference metric, the DSC and the CD metric as defined in section 2.2 and qualitatively by visual inspection.

Table 2.1 and 2.2 contains relevant clinical information for the 8 patients used as cohorts for this study.

Item	P1	P2	P3	P4
Diagnosis	Malignant neoplasm of bronchus or lung	Malignant neoplasm of breast	Overlapping malignant neoplasm of lip, oral cavity and pharynx	Malignant neoplasm of central nervous system
Volume site	Right Lung	Right Breast + Right SCF	Head and Neck - Floor of mouth, L Ant Neck, R Ant Neck	Craniospinal + post fossa boost
Treatment type	3DCRT	3DCRT	IMRT	3DCRT
Prescibed dose (Gy)	50	50, 50	60, 50, 50	55.8
Number of fractions	20	25, 25	30, 25, 25	31
Date of initial pCT	08/10/2013	20/08/2013	16/10/2013	11/10/2013
Date of first fraction	08/10/2013	21/10/2013	18/11/2013	29/10/2013
Date of re-CT	21/10/2013	08/10/2013	15/11/2013	08/11/2013
Reason for re-CT	Reduction in collapsed lung	Implant removed from her left breast	Mass increased in size - swelling on the mandible	Setup was difficult to reproduce
Time between pCT and re-CT (days)	13	49	30	28
Time between first fraction and re-CT (days)	13	N.A.	N.A.	10

Table 2.1 Relevant Clinical Information I

Item	P5	P6	P7	P8
Diagnosis	Overlapping malignant neoplasm of tonsil	Malignant neoplasm of parotid gland	Malignant neoplasm of prostate	Malignant neoplasm of vulva
Volume site	Head and Neck - R Tonsil, R Ant Neck, L Ant Neck	Head and Neck - Right Parotid	Prostate	Groin
Treatment type	IMRT	Phase 1:3DCRT - Phase 2:IMRT	VMAT	VMAT
Prescibed dose (Gy)	75, 50, 50	27.5, 27.5	74	50
Number of fractions	35, 25, 25	10, 10	37	25
Date of initial pCT	29/11/2013	15/11/2013	14/10/2013	13/12/2013
Date of first Fraction	19/12/2013	25/11/2013	04/11/2013	02/01/2014
Date of re-CT	29/01/2014	16/12/2013	18/11/2013	15/01/2014
Reason for re-CT	4.4 kg weight loss (5% was ~88kg) needed to assess spinal cord tolerence	Two phase plan	Rectum constantly in different position to CT	Swelling in the nodal region
Time between pCT and re-CT (days)	61	31	35	33
Time between first fraction and re-CT	41	21	14	13

Table 2.2 Relevant Clinical Information II

2.4.2 Head and neck study: without user input

For this retrospective study a cohort of 6 patients were selected from the Pinnacle TPS database and imported into RayStation.

The 6 patients selected for this study were head and neck cases. A single treatment site was chosen for the purpose of presenting the algorithm with multiple CT image data sets which are relatively similar, this helps assess how well the algorithm deforms for this particular site and how consistently it applies the deformation.

The head and neck anatomical site does not typically deform as much as other sites, such as the thoracic and pelvic region. This is because the structures within the head and neck region are typically more stable, particularly the bony anatomy such as the skull and the soft tissue anatomy such as the brain. This means that the head and neck anatomical site serves as a good basis for the validation and verification of the DIR algorithm as it should in theory be a relatively simple site.

A range of contours were delineated in the head and neck region, each contour studied was delineated on a minimum of two patients for comparison. There were 42 contours assessed in this study, the contours included were the spinal cord, spinal cord PTV, brainstem, brainstem PTV, mandible (including left and right), chiasm, left optic nerve, left optic nerve PRV, and the brachial plexus (left and right).

All contours were deformed using the DIR algorithm and according to the DIR method in section 2.1. The contours were assessed quantitatively using the DVR and the DSC as defined in section 2.2, and qualitatively by visual inspection. The propagated contours were also assessed on RayStation using a simple optimised plan (SOP), that is, a dose distribution of 70 Gy in 35 fractions to the tumour was planned and optimised around the propagated

contours on the target CT image set for each patient. If the propagated contours were incorrectly deformed then it is expected that the PTV will be underdosed and the surrounding tissue will be overdosed, with OARs likely to be overdosed. The difference in the dose to the propagated contours and the expertly defined contours on the target CT image set will be assessed to determine the dosimetric implications of using propagated contours for clinical planning.

ICRU Reports 50, 62 and 83 recommend that the dose delivered to the target volume in the patient is within -5% and +7% of the prescribed dose, this underpins the principle used in treatment planning which is to ensure that 100% of the volume receives at least 95% of the prescribed dose. There are a number of qualitative dose statistics that are used to report the dose to the target volume and the OAR. The dose statistics calculated and reported in this study are ones recommended by ICRU Reports 50, 62 and 83, they are the $D_{98\%}$, the $D_{2\%}$, the D_{mean} and the D_{max}. The $D_{98\%}$ is the maximum dose received by at least 98% of the volume and is useful for determining dose coverage to the PTV, the $D_{2\%}$ is the maximum dose received by at least 2% of the volume, the D_{mean} is the mean dose to the volume and is useful for measuring the dose uniformity to the PTV, the D_{max} is the maximum dose to the volume and is useful for looking at the maximum tolerance dose to OAR. The maximum dose to an OAR is particularly important for serial structures, as when a part is irradiated above tolerance dose the OAR stops functioning completely, this is in contrast to parallel organs which can have parts of the organ exceed tolerance dose but still function at a lower capacity [16].

2.4.3 Head and neck study: with user input

The DIR algorithm in RayStation allows the user to choose ROIs that are contoured on both the reference CT data set and the target CT data set to be controlling structures and focus structures.

The controlling structure function is designed to guide the deformation algorithm. A penalty term is added to the objective function that drives the registration, as described in section 1.2, and aims to deform the contours from the reference CT image data to the corresponding contours on the target CT image data set [29].

The focus structure function is designed to calculate a deformation in a specific volume of the CT data set, as determined by the ROI selected, outside of which the algorithm does not calculate any deformation. By selecting an ROI as a focus ROI, the volume surrounding it is the specific volume that is "focused" on by the algorithm [29].

The two functions were used on 15 selected contours from the head and neck anatomical site study. The DIR algorithm was implemented according to the DIR method in section 2.1 with the exception that the appropriate controlling structure or focus structure was also selected. A separate DVF was therefore calculated for the DIR alone (i.e. without a controlling or focus structure), the DIR with a controlling structure and the DIR with a focus structure for each of the selected contours. The contours were then propagated and assessed qualitatively using the DSC metric as defined in section 1.2. This meant that there were 15 DSC values for the DIR algorithm alone, 15 DSC values for the DIR algorithm with a controlling structure and 15 DSC values for the DIR algorithm with a focus structure. The DSC values for the DIR algorithm alone was used to compare to the DSC values with a controlling structure, and with a focus structure. A paired, two-tailed t-test was used to see if there was any statistically significant difference between the DSC values.

2.5 Dose Measurements

2.5.1 Phantom Layers

A relatively simple phantom arrangement was used to simulate a change in anatomy for the purpose of dose propagation and dose measurements. Ten layers were placed on top of a water phantom which has a central ion-chamber insert. Nine layers were made from solid plastic polystyrene, and one layer was made of cork. A single layer of polystyrene was placed on top of the phantom, the next layer was cork and the remaining eight layers of polystyrene were placed on top of that.

Polystyrene was used as it is a water equivalent phantom material and is therefore dosimetrically equivalent to soft tissue, with a density of 1.05 g/cm^3 [7]. Cork was used as it is dosimetrically equivalent to lung tissue [42]. These properties made the materials ideal for anatomical tissue change simulation. The water phantom and layers were 20 x 20 cm in length and width. The water phantom was 11 cm high, the polystyrene layers were 0.31 cm thick and the cork layer was 0.6 cm thick. Figure 2.3 illustrates the arrangement.

This phantom arrangement enabled the validation of the DIR algorithm using dose measurements. The phantom was scanned on a Siemens sensation open CT. The phantom was scanned in two arrangements, one with all ten layers placed on top of the phantom, and one with seven layers of polystyrene with the cork layer and two polystyrene layers removed. The two scans were used to simulate an expansion and contraction of tissue. The seven layer phantom was set as the reference for the simulated expansion and the ten layer phantom was set as the reference for the simulated contraction.

A static 10 x 10, 6X beam was planned on the target image set in RayStation to deliver

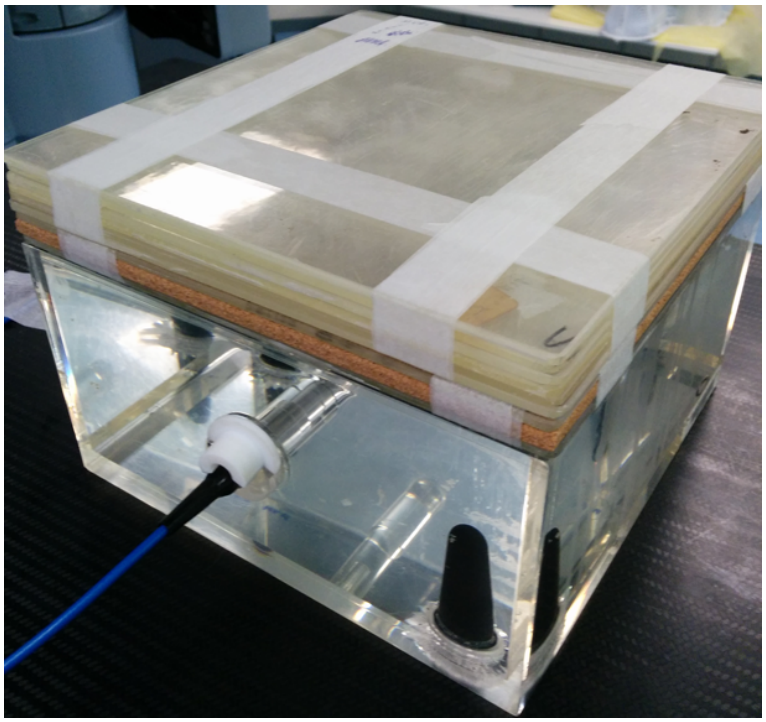


Fig. 2.3 Dose phantom arrangement

120 monitor units (MUs) with the isocentre in the middle of the ion-chamber. A DVF was calculated from the reference to the target and the dose calculated for the 10 x 10 field was deformed from the target to the reference image set. The beam was delivered on an iX Linear accelerator. The dose was measured using a 10 x 10 cm piece of gafchromic film placed centrally on the top surface of the photon water phantom, this gave a measured dose in this plane. A point dose measurement was also made using an ion-chamber at 5 cm depth in the photon water phantom.

The dose measurements were made on the phantom arrangement that was scanned for the reference image set, as the dose is deformed from the target to the reference image set. The dose that the ion-chamber and film was expected to receive was calculated using DIR with the deform dose function. The dose that the ion-chamber and film were expected to receive was also calculated on RayStation by planning the same beam arrangement, including energy and MUs, as was on the target image set.

Chapter 3

Results and Discussion

3.1 Phantom Study

3.1.1 Multiple layer solid water phantom: without user input

The mean DSC for the expansion DIR, contraction DIR and the rigid registration was 0.93 +/- 0.03, 0.98 +/- 0.01 and 0.96 +/- 0.01 respectively. A paired, two-tailed t-test showed that there is a statistically significant difference between the expansion and contraction DSC values (p=0.01). A paired, two-tailed t-test showed that there wasn't statistically significant difference between the expansion DIR and rigid registration DSC values (p=0.14). A paired, two-tailed t-test showed that there is a statistically significant difference between the contraction DIR and rigid registration DSC values (p=0.01). The results indicate that using DIR for tissue expansions and contractions is at least as good as rigid registration.

The DSC results indicate that the simulated contraction was more accurate than the simulated expansion, therefore the DIR algorithm may be more accurate at deforming contract-

ing structures as opposed to expanding structures. If this were to translate into real patient image data, then the algorithm may be better suited for tissue loss as opposed to tissue gain, for example in tumour reduction as opposed to tumour growth. The results need to be taken in consideration with the understanding that the DSC is an indicator of the registration accuracy, and that it can not be relied on solely for contour evaluation purposes, thorough user evaluation is still required.

The POIs were placed in the plane approximately 2 mm below the top surface and above bottom surface indicated as plane 1 and 2 in figure 3.1 and figure 3.2 (coronal image 58th slice of both the expansion and contraction). It's reasonable to expect a calculated shift of approximately zero for the x and y directions and an absolute shift close to 3 mm in the z-direction for both the simulated contraction and expansion in plane 1 which is near the top surface of the phantom. It is also reasonable to expect a calculated shift of approximately zero for the x, y and z coordinates for both the simulated contraction and expansion in plane 2 which is near the bottom surface of the phantom.

The average POI shift in cm for the simulated expansion and contraction in tissue over the four phantom arrangements in the planes 1 and 2 are listed in table 3.1, figures 3.3, 3.4, 3.5 and 3.6 show a transverse view of these planes. For the figures 3.1 to 3.6 the colour blue represents vectors 0 to 0.10 cm in length, yellow represents vectors 0.10 to 0.15 cm in length, green represents vectors 0.15 to 0.25 cm in length and red represents vectors 0.25 to 0.30 cm in length.

	Simulated contraction TRE (cm +/- 0.1)	Simulated expansion TRE (cm +/- 0.1)
Plane 1	(-0.01, 0.02, -0.29)	(-0.01, -0.01, 0.08)
Plane 2	(-0.01, 0.02, -0.01)	(-0.01, 0.00, 0.18)

Table 3.1 POI results for multiple layer solid water phantom: without user input

The POI shifts for the simulated contraction were closer to what could be reasonably expected than the simulated expansion. The simulated contraction had a shift of <0.2 mm for the x and y directions in both planes and the z direction in plane 2, and a shift of approximately 2.9 mm in the z-direction in plane 1. The simulated expansion also had a shift of <0.2 mm in the x and y directions for both planes, however the z-direction was approximately 0.8 mm and 0.18 mm for planes 1 and 2 respectively which is outside the expected values. From these results it is likely the algorithm is calculating a more accurate DVF for the simulated contraction than the simulated expansion. The POI results also indicate that for this phantom arrangement the algorithm is calculating a better DVF transform for the simulated contraction compared to the simulated expansion which is registered in the opposite direction. This indicates that the algorithm may be better at deforming tissue contractions as opposed to tissue expansions.

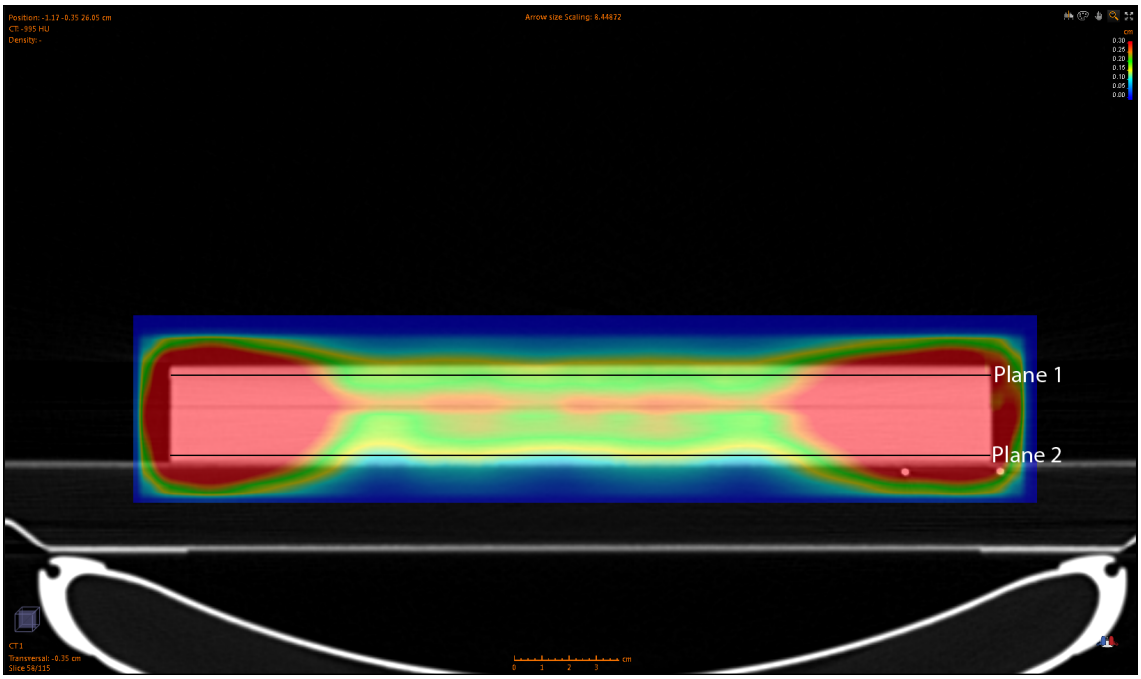


Fig. 3.1 Simulated Expansion (coronal)

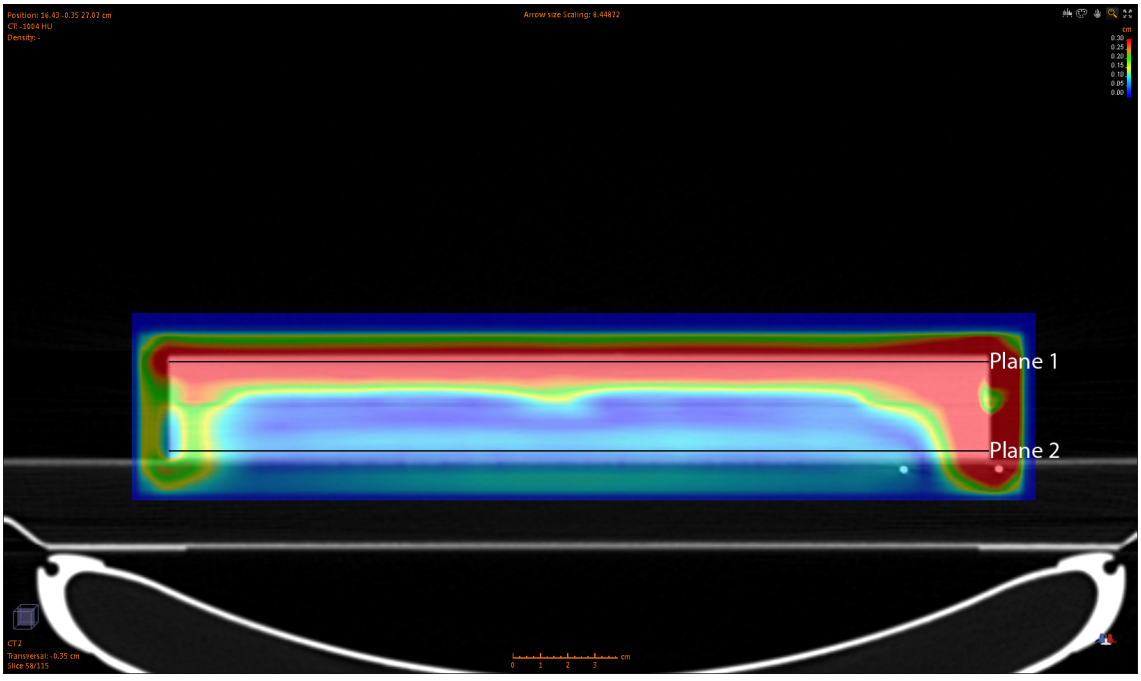


Fig. 3.2 Simulated Contraction (coronal)

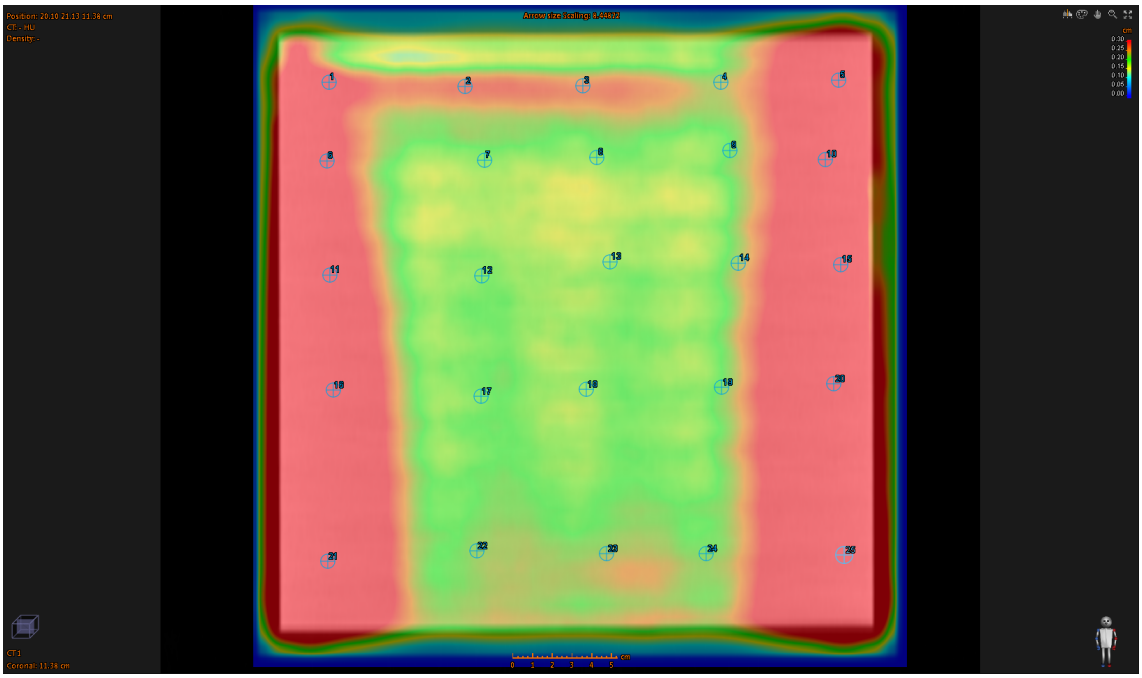


Fig. 3.3 Simulated Expansion - Plane 1 (transverse)

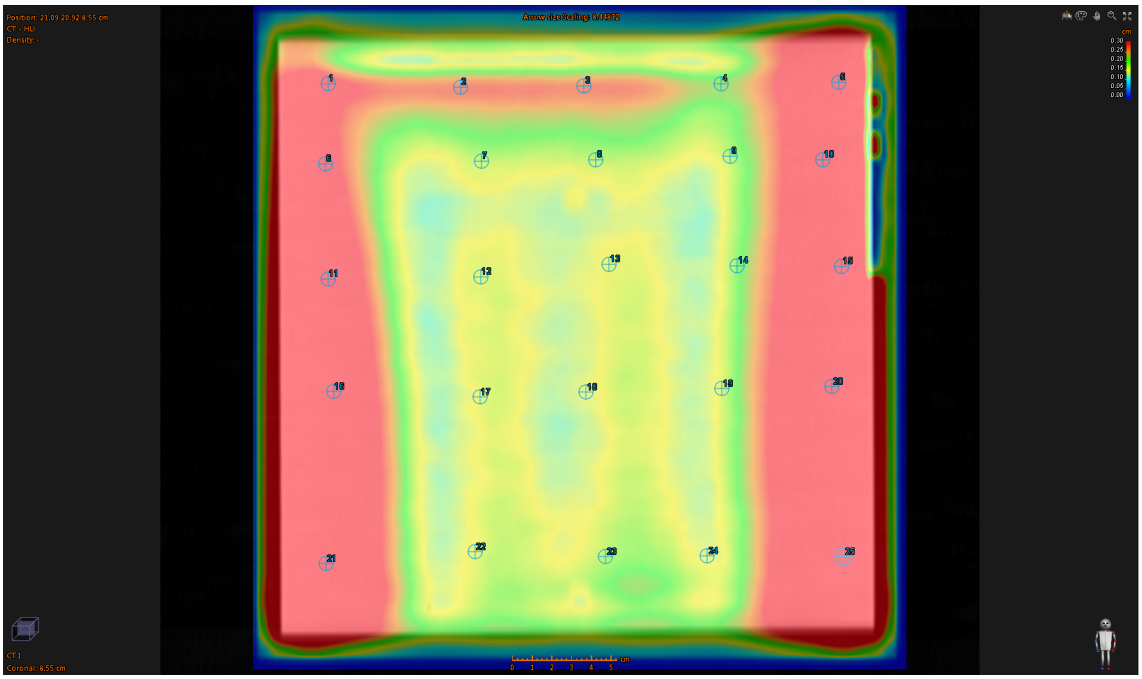


Fig. 3.4 Simulated Expansion - Plane 2 (transverse)

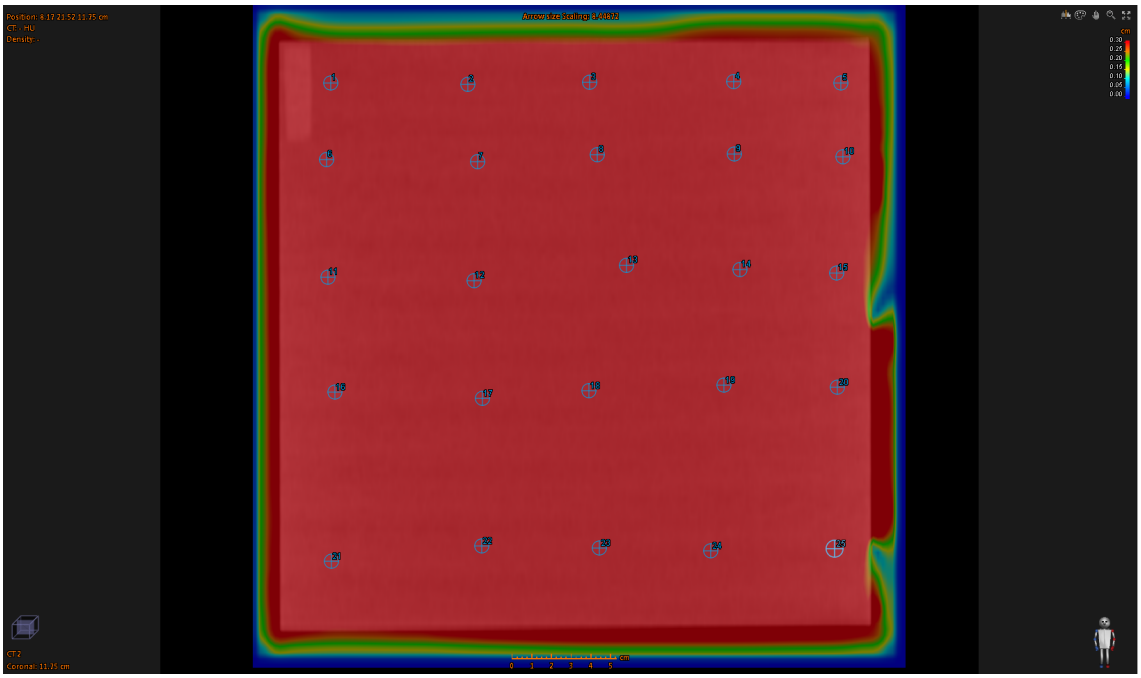


Fig. 3.5 Simulated Contraction - Plane 1 (tansverse)

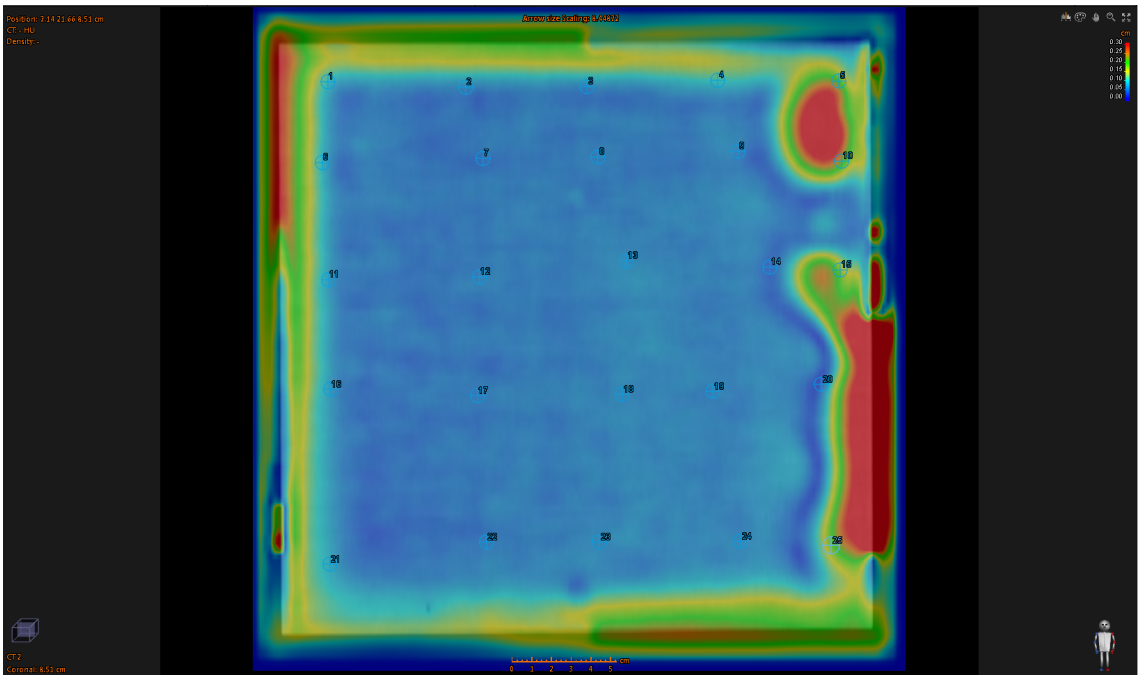


Fig. 3.6 Simulated Contraction - Plane 2 (transverse)

Visual inspection was used to qualitatively assess the calculated DVFs. For the simulated contraction and expansion the colour blue represents vectors 0 to 0.12 cm in length, green represents vectors 0.12 to 0.23 cm in length, yellow represents vectors 0.23 to 0.29 cm in length and red represents vectors 0.29 to 0.35 cm in length. The DVF calculated for the simulated contraction appeared to be superior than the DVF calculated for the simulated expansion as can be seen in figure 3.1, 3.2, 4.1 and 4.2.

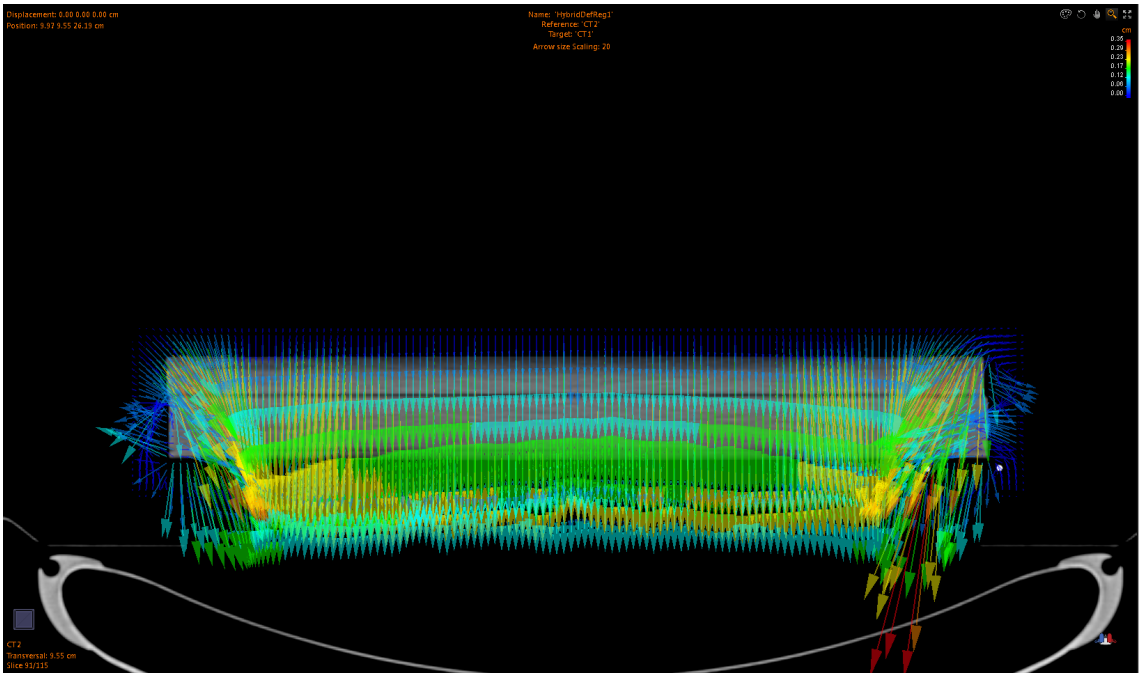


Fig. 3.7 Simulated Contraction DVF

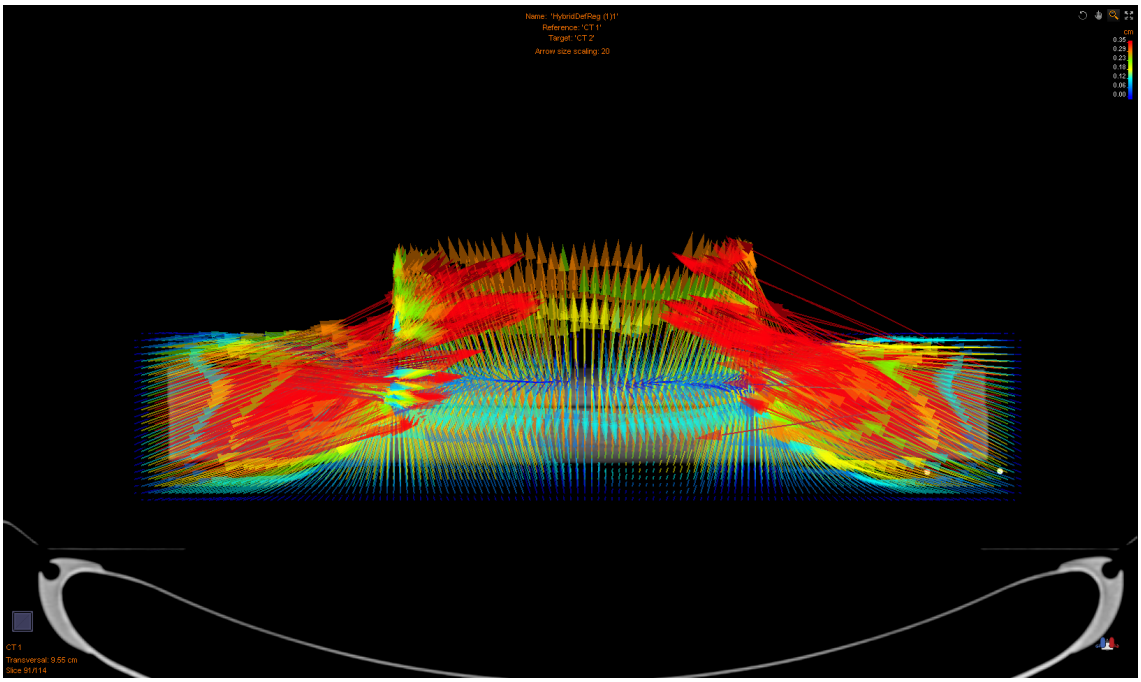


Fig. 3.8 Simulated Expansion DVF

From visual inspection the vectors point in the expected direction for both the expansion and contraction, which is upwards and downwards respectively. However both DVFs displayed incorrect vectors of at least 3 mm in length in locations along the edges of the phantom, with this being more pronounced in the expansion simulation. Most of the vectors still have a component in the correct direction, however they also have relatively large components in incorrect directions which results in the vectors pointing off to an angle. It is clear that the vectors were more uniform in the contraction simulation compared to the expansion simulation. All DVFs were relatively uniform towards the centre of the phantom, and relatively non-uniform towards the edges. This indicates that the algorithm did not deform the edges of the phantom as accurately as the central region.

The qualitative and quantitative results show that the algorithm is calculating a better DVF for the simulated contraction than the simulated expansion. This indicates that the algorithm handles the change in CT values from 1000 HU to 0 HU better than the opposite

direction. This has implications in the use of the algorithm on patients, as weight loss in the head and neck is likely to be better determined by the algorithm and reflected in the DVF than swelling in the head and neck for example. Or more gas in the rectum in the follow up scan could not be accounted for as well by the algorithm as less gas in the rectum. It is possible the algorithm sees the problem as "missing" CT numbers when removing the 270cm^3 volume and going from 1000HU to 0HU between the reference and the target CT image data sets. Therefore the algorithm may handle "missing" CT numbers better than the reverse, i.e. "creating" CT numbers.

The results indicate that the contours have propagated incorrectly for both the simulated expansion and contraction. Incorrect contour propagation has potential dosimetric implications for treatment. For example, if the propagated contours are accepted for treatment planning as a part of ART then there could be serious dosimetric consequences to the target volume and OARs, with underdosing to the target tumour and overdosing to the OARs being likely outcomes, which subsequently affects the tumor control probability and tissue complications probability and therefore the therapeutic ratio. The results indicate that there is a difference in the ability of the DIR algorithm to deform a simulated expansion of tissue as compared to a simulated contraction of tissue, with the simulated contraction being more accurate. However the simulated contraction is not ideal either, it would therefore be absolutely necessary to have an expert user thoroughly evaluate any propagated contour that is intended to be used for clinical treatment planning purposes, this result is also found in a number of research papers [36] [38].

A limitation of this setup is that the phantom arrangement is not an ideal situation and does not reflect a clinical scenario as well as it potentially could. Although the CT numbers are similar to the tissue density of a real subject, the shape of the square layers is not realistic. However the it does give an indication of the limitations of the algorithm.

Conclusion - Multiple layer solid water phantom: without user input

The results indicate that the DIR algorithm was more accurate when calculating a DVF for the simulated contraction of tissue, as opposed to the simulated expansion of tissue. Therefore the DIR algorithm may be more accurate at deforming tissue contractions as opposed to tissue expansions with real patient CT image sets, for example weight loss as opposed to weight gain and tumour reduction as opposed to tumour growth. This has important implications for contour propagation, as any contour that is propagated in a region of tissue expansion may be less accurate than a contour propagated in a region of tissue contraction.

Accepting propagated contours for clinical treatment planning requires careful evaluation, if the contours are not evaluated by an expert user then there is a high possibility of negative dosimetric consequences with underdosing to the target tumour and overdosing to the OARs being possible outcomes. The results indicate that this is particularly important for contours propagated in regions of tissue expansion, however the results also indicate that all propagated contours have errors associated with them, and so it is important that all propagated contours are thoroughly evaluated and edited by an expert user if they are to be used clinically for treatment planning or dose accumulation purposes.

3.1.2 Multiple layer solid water phantom: with user input

Visual inspection was used to evaluate the deformed images. The deformed image is the reference image with the deformations applied and therefore ideally it should be identical to the target image. The deformed image provides a good visual tool for inspecting how well the algorithm is deforming the image. From visual inspection of the deformed images, including figure 3.9 and 3.10, the deformation appeared to be more accurate with the

controlling structure than without for both the contraction and expansion simulation.

For the simulated contraction and expansion there was a distortion on the edges of the phantom that erroneously extended part the phantom up to approximately half a cm out from the edge. On some of the CT image data sets the CT couch merged into the phantom as shown in figure 3.9. The use of the external contour as a controlling structure significantly decreased the distortion on the edges of the phantom as shown in figure 3.10, the edges became essentially parallel with a relatively small amount of distortion.

The results indicate that the use of the external contour as a controlling structure helps guide the DIR algorithm to achieve more accurate DVFs. The DVF could be used to propagate contours or dose, it is therefore important that it is as accurate as possible. However there is still a significant error associated with DIR calculations and it is therefore necessary to have an expert user thoroughly evaluate the DVF by visual inspection of the DVF itself as well as any propagated contours if it is to be used clinically.

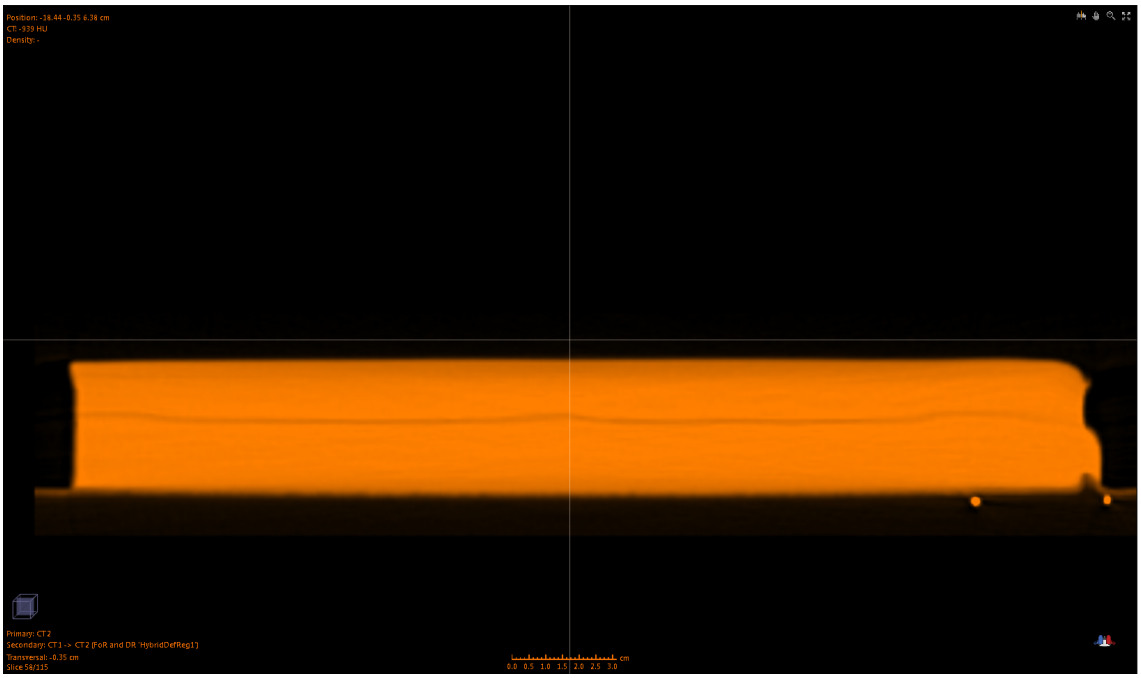


Fig. 3.9 Solid water phantom with no controlling structure

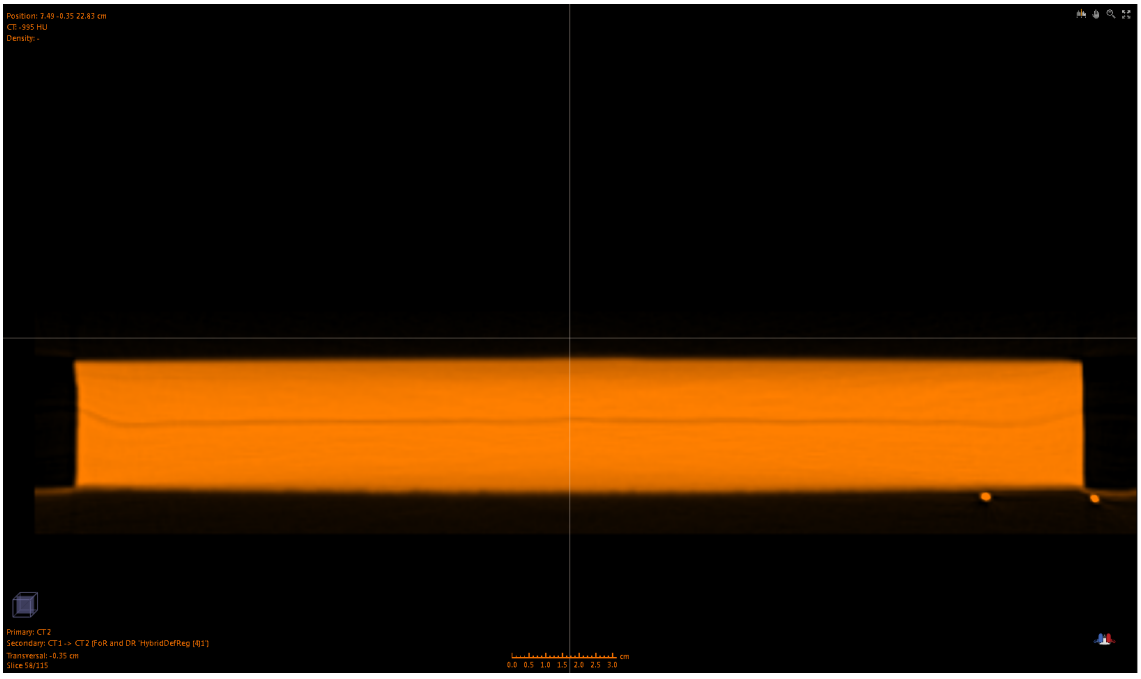


Fig. 3.10 Solid water phantom with the external contour selected as a controlling structure

Conclusion - Multiple layer solid water phantom: with user input

From visual inspection it was found that using the external contour as a controlling structure increases the accuracy of the deformation algorithm. The influence of the CT couch on the algorithms calculation was minimised by using the external as a controlling structure. Using the external contour as a controlling structure is therefore a relatively easy and effective method for improving the algorithm's accuracy.

An external contour should be used as a controlling structure for any DIR calculation, as it helps improve the accuracy of the DVF. However, there are still errors associated with the DVF, therefore every DVF needs to be visually assessed by an expert user and any contours propagated should also be visually assessed by an expert user. Similarly, any dose propagated needs to have an accurate DVF as determined by an expert user.

3.1.3 Effect of image quality on DIR

Table 3.2 gives the image quality information for each scan type used, as measured by the CATPHAN 600. Standard deviation is used to quantify noise, a smaller standard deviation represents a higher image quality. The standard deviation of the CT scans with the lowest, normal and highest image quality was 30, 13 and 8 respectively. The CBCT had a standard deviation of 14. As expected the standard deviation decreased with decreasing noise. The contrast resolution was also affected, with the contrast resolution increasing with increasing image quality. The resolution however was the same across all CT scans and slightly better for the CBCT scan.

	Image Noise (Mean HU +/- 1 S.D.)	Contrast Resolution	Resolution (line pairs/cm)
CT: Lowest image quality	-13 +/- 30	15 mm sphere on the 1% supra-slice	4
CT: Normal image quality	13 +/- 13	7 mm sphere on the 1% supra-slice	4
CT: Highest image quality	20 +/- 8	6 mm sphere on the 0.3% supra-slice	4
CBCT: Pelvis scan	4 +/- 14	15 mm sphere on the 1% supra-slice	5

Table 3.2 Image quality parameters

Solid water phantom: image quality assessment

The DSC results are listed in table 3.3. The difference between the simulated contraction and expansion was not statistically significant (p = 0.31).

	Simulated contraction DSC	Simulated expansion DSC
Lowest image quality	0.98	0.97
Normal image quality	0.98	0.97
Highest image quality	0.98	0.86

Table 3.3 DSC results for solid water phantom: image quality assessment

The deformed images of the simulated contraction are shown in figure 3.11, 3.12 and 3.13, these are a single transverse slice close to the centre of the phantom which is representative of the other slices as the deformed phantoms were relatively uniform. The reference contour is the solid line and is the contour that outlines the phantom in the reference image set, the target contour is the contour that outlines the phantom in the target image set and is the dashed contour line. Ideally the phantom should be contained within the target contour when visually assessing the deformed image.

From visual inspection of figures 3.11, 3.12 and 3.13, it was observed that the left side of the phantom extended out beyond the edge boundary by approximately 0.6 cm, 0.5 cm and 0 cm for the lowest, normal and highest image quality respectively. However for the highest image quality the left side of the phantom was consistent with the the edge boundary. This indicated that the deformation accuracy increased with increasing image quality, however the highest image quality did not completely fill the target contour which represents the target phantom and was therefore not entirely accurate as also indicated by the DSC values. The deformed images of the simulated contraction also show that for all three image qualities the phantom did not move significantly from the reference position, outlined by the solid contour line, to the target position, outlined by the dashed contour line.

The deformed images of the simulated expansion are shown in figure 3.14, 3.15 and 3.16, these are a single transverse slice close to the centre of the phantom. From visual inspection it was observed that parts of the left and right sides of the phantom extended beyond the edge boundary by approximately 0.6 cm for the lowest image quality. For the normal image quality a small part of the right side extended approximately 0.4 cm beyond the edge of the phantom and parts of the top surface extended to the target dashed contour. For the highest image quality the left side extended beyond the edge of the phantom by approximately 0.5 cm, the anterior surface extended beyond the target dashed contour by approximately 0.25 cm and the posterior surface by approximately 0.3 cm. For the simulated expansion the DSC increased from the lowest image quality to the normal image quality, but decreased for the highest image quality, this was not expected but was consistent with visual inspection.

From visual inspection the extension of the phantom beyond the edge of the phantom on the left and right sides represents erroneous calculations by the deformation algorithm as these are locations where no deformation has been simulated and hence zero deformation

should be observed. The results need to be taken in consideration with the limitations of the method. It has been suggested that a possible limitation is the uniformity of the phantom, clinical images contain intensity gradients however the phantom is of uniform intensity which is anatomically unrealistic and provides the algorithm with information it would not see clinically [42]. The deformation applied to the phantom is also anatomically unrealistic. However it is still a useful simulation and does test the limits of the algorithm.

This arrangement has provided results that indicate image quality does influence the calculations of the DVF by the DIR algorithm being assessed. The results indicate that better image quality is more accurate for DIR in regions of simulated contraction and that normal image quality is more accurate than other image qualities for DIR in regions of simulated expansion. Both the expansion and contraction of tissue occurs in a real patient, the results suggest that using a higher image quality does not significantly increase the accuracy of DIR for both expansion and contraction of tissue and is in fact worse for the contraction of tissue, it would therefore not be justified to increase the image quality and associated imaging dose according to the justification and optimisation principles recommended in the ICRP-60 publication [24].

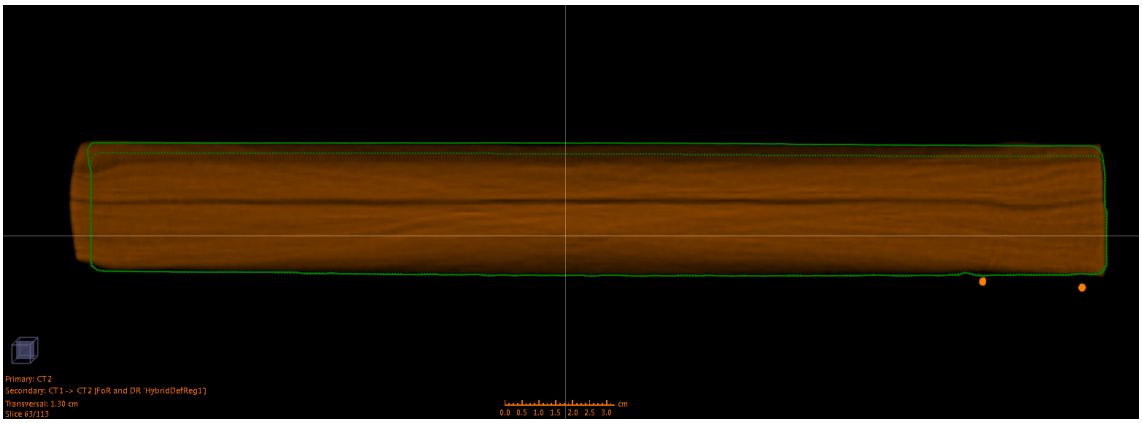


Fig. 3.11 Lowest image quality - Simulated contraction (coronal)

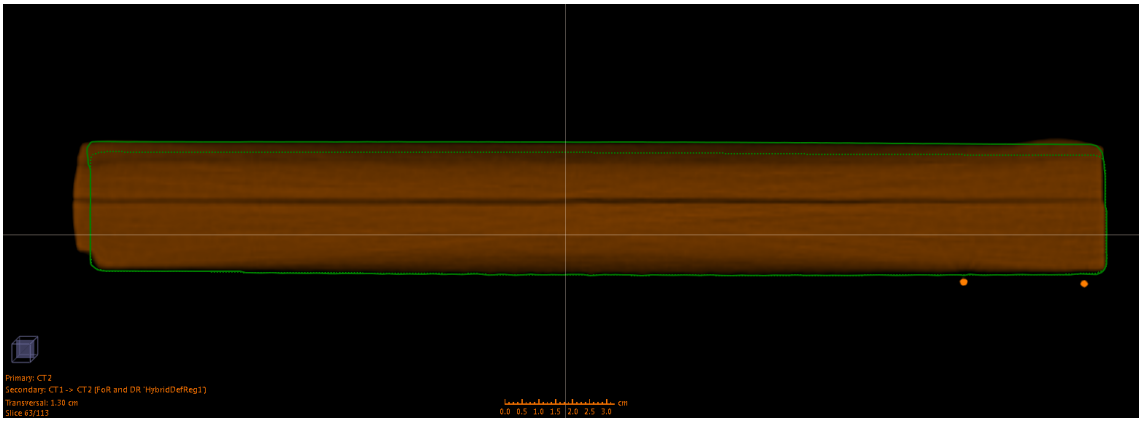


Fig. 3.12 Normal image quality - Simulated contraction (coronal)

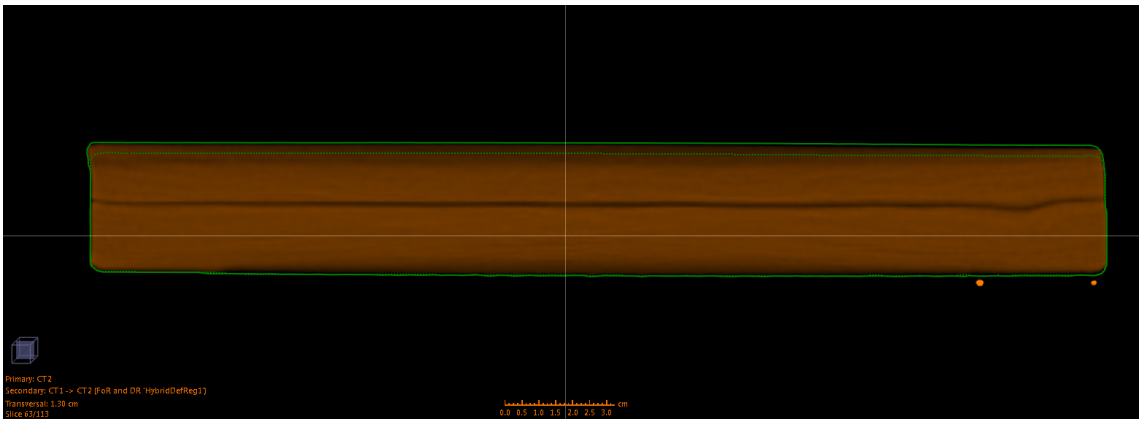


Fig. 3.13 Highest image quality - Simulated contraction (coronal)

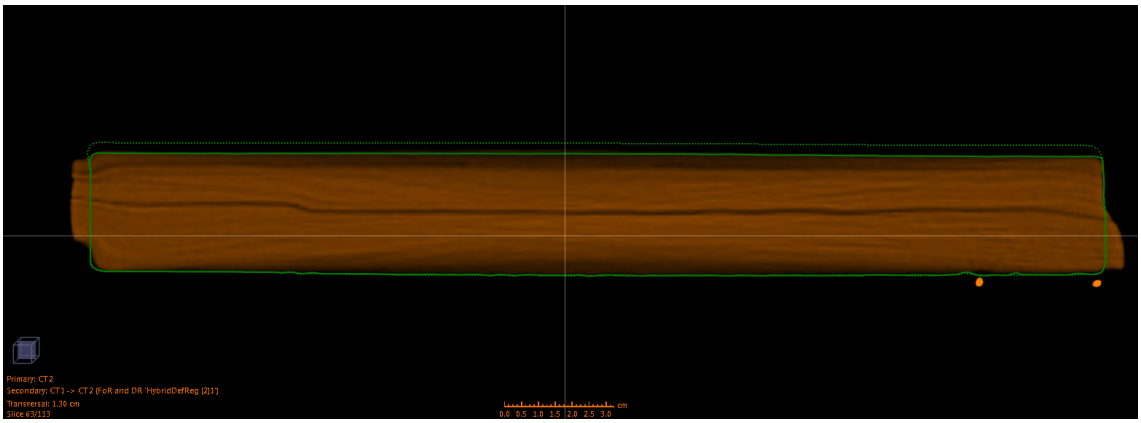


Fig. 3.14 Lowest image quality - Simulated expansion (coronal)

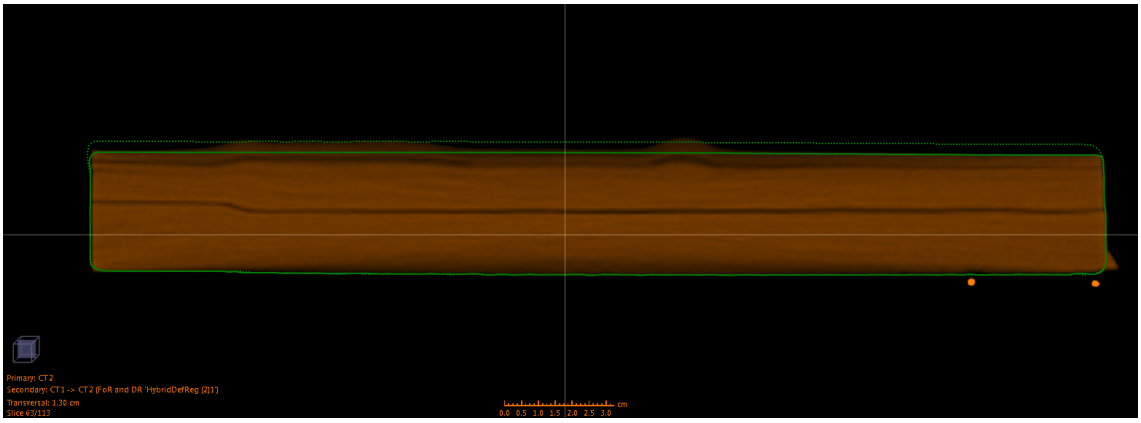


Fig. 3.15 Normal image quality - Simulated expansion (coronal)

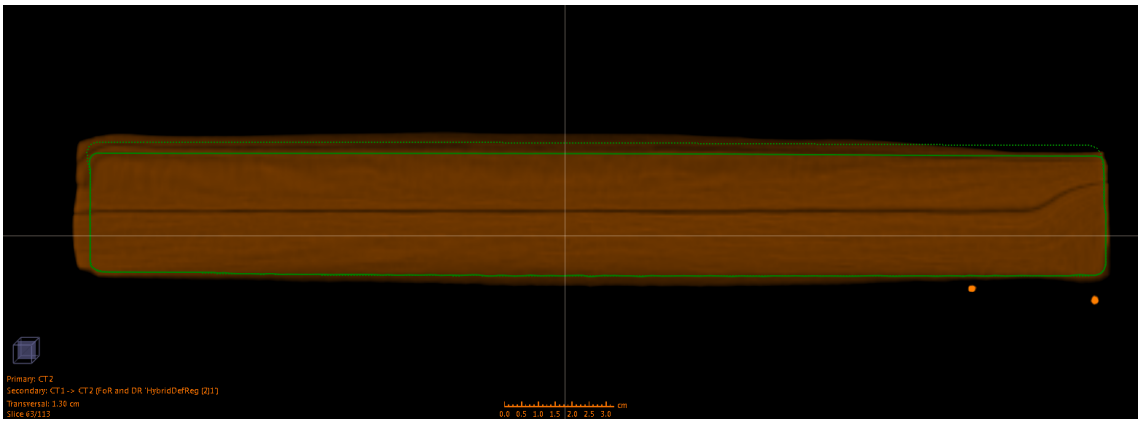


Fig. 3.16 Highest image quality - Simulated expansion (coronal)

Conclusion - Solid water phantom: image quality assessment

The results indicate that the DIR algorithm is more accurate at calculating a DVF for the simulated contraction with a higher image quality than lower image qualities, however the results also indicate that the DIR algorithm is most accurate for the simulated expansion with the normal image quality. The results indicate that increasing the image quality would not significantly improve the accuracy of the DVFs calculated by the DIR algorithm in a real patient, therefore increasing the image quality and the associated imaging dose is not in line with the justification and optimisation principles recommended by ICRP-60 publication [24]. Increasing the imaging dose is not justified as the benefit does not outweigh the cost.

Apple phantom: image quality assessment

The DSC was calculated for each registration of the CT to a different CBCT image quality. The DSC results are listed in table 3.4. The mean DSC for the simulated contraction and simulated expansion for all image qualities was 0.88 +/- 0.00 and 0.90 +/- 0.01. A paired, two tailed t-test showed that there wasn't a significant difference between the mean simulated contraction and expansion DSC (p=0.06). The DSC also had a relatively small standard deviation between image qualities, indicating that the image quality wasn't a significant factor in the registration. From these results the higher image quality would not be justified due to the increased dose delivered to the patient.

	Simulated contraction DSC	Simulated expansion DSC
Lowest image quality	0.89	0.90
Normal image quality	0.88	0.90
Highest image quality	0.88	0.90

Table 3.4 DSC results for apple phantom: image quality assessment

The TRE was calculated for 10 points identified on the surface of the apple, in both the reference and target image sets. The same points were identified on all registrations so that they could be compared. Figure 3.17 shows the identified points on a single slice of the normal image quality CT to CBCT registration for both the simulated expansion (right) and simulated contraction (left). The points identified were identical for each of the different image qualities used. Left-Right is the x-direction, Anterior-Posterior is the y-direction and Superior-Inferior is the z-direction. The colour blue represents vectors 0 to 0.33 cm in length, green represents 0.33 cm to 0.67 cm in length, yellow represents vectors 0.67 to 0.84cm in length and red represents vectors 0.84 to 1cm in length.

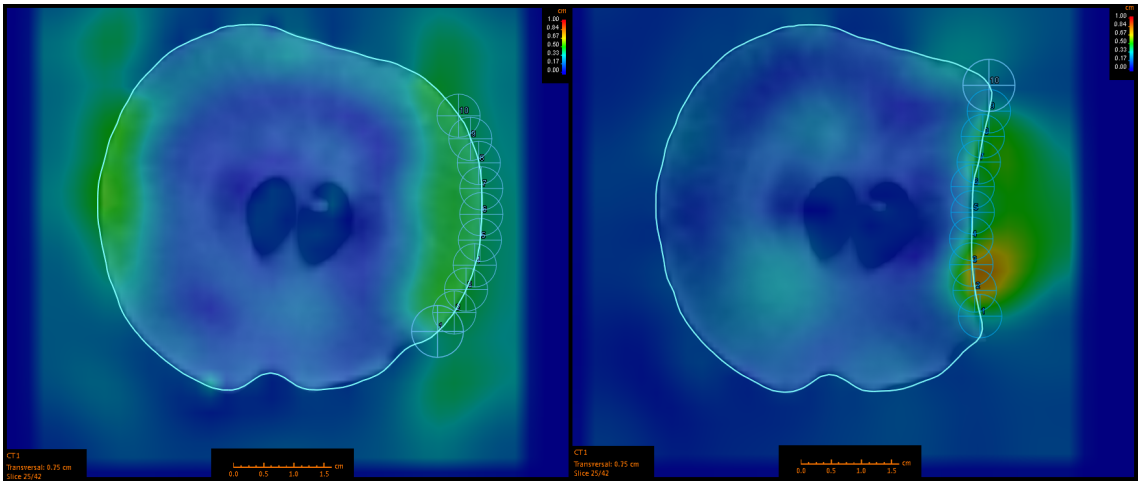


Fig. 3.17 (Left) simulated contraction (Right) simulated expansion

The TRE results are listed in table 3.5. Ideally the TRE would be zero which would indicate that the POIs deformed from a number of identified points relative to the anatomy on the reference image to the same points relative to the anatomy in the target image, any deviation from the identified points will results in a non-zero TRE. The TRE is relatively large for the x-coordinate, with between 0.4 and 0.5 cm error from the ideal position for the simulated expansion and between 1.2 and 1.3 cm error for the simulated contraction. This indicates that the algorithm has not deformed the apple accurately. When considering

either the simulated expansion or contraction the accuracy of the algorithm as indicated by the TRE does not change significantly for different image qualities. This indicates that the image quality does not have a significant influence on the algorithms calculations.

	Simulated contraction TRE (cm +/- 0.1)	Simulated expansion TRE (cm +/- 0.1)
Lowest image quality	(-1.2, -0.1, -0.0)	(-0.4, -0.2, 0.0)
Normal image quality	(-1.2, -0.1, -0.0)	(-0.5, -0.1, 0.0)
Highest image quality	(-1.3, -0.1, -0.1)	(-0.5, -0.2, 0.1)

Table 3.5 TRE results for apple phantom: image quality assessment

Three transverse images of the apple on the same slice (0.75 cm), for all CT to CBCT simulated expansions are shown in figure 3.18 and simulated contractions are shown in figure 3.19. From visual inspection it is evident that the DIR algorithm did not fully deform the contour from the reference to the target image for both the simulated expansion and contraction. The DIR algorithm did deform the contour in the correct direction and identified deformation in the deformed region, however the magnitude of the deformation was not calculated accurately by the DIR algorithm. Ideally the blue dashed line which represents the propagated contour and the magenta dashed line which represents the expertly defined target contour would be identical. For the simulated expansion, the algorithm incorrectly deformed the contour on the left side of the apple, which decreased the size of the contour. In a real patient an incorrect decrease of a contour on a target volume would result in an underdose, which could have significant implications for treatment.

The quantitative and qualitative results indicate that the DVF has errors for all image qualities and that every DVF calculated would need to be thoroughly evaluated by an expert user if it was to be used clinically for contour or dose propagation, and that any contours propagated for clinical use would also have to be thoroughly evaluated by an expert user.

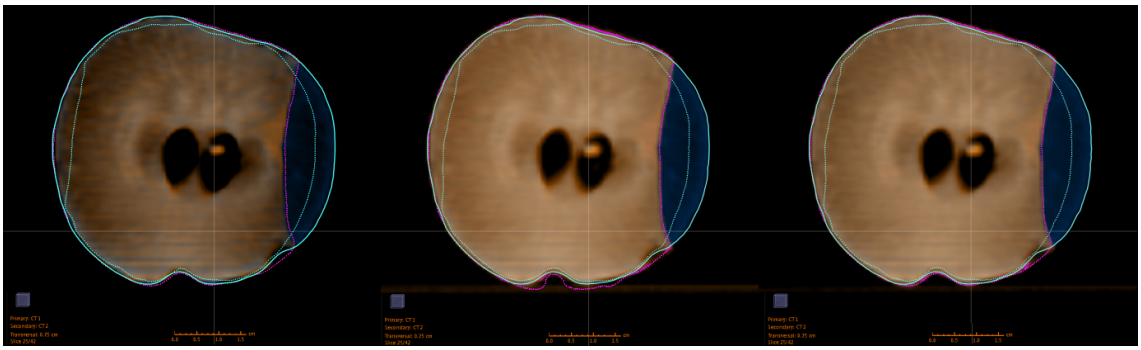


Fig. 3.18 Simulated Expansion, magenta dashed is the target external contour, blue solid is the reference external contour and blue dashed is the propagated contour, ideally the two dashed lines would overlap. (Left) lowest CT image quality, (Centre) Normal CT image quality and (Right) Highest CT image quality.

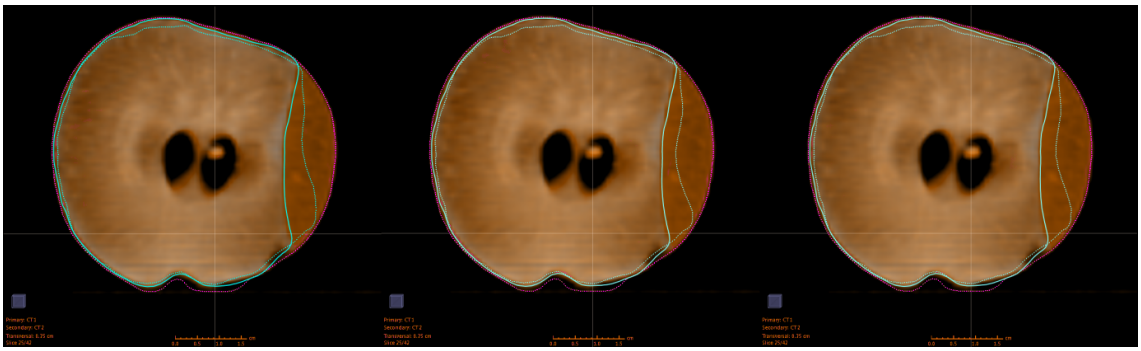


Fig. 3.19 Simulated Contraction, magenta dashed is the target external contour, blue solid is the reference external contour and blue dashed is the propagated contour, ideally the two dashed lines would overlap. (Left) lowest CT image quality, (Centre) Normal CT image quality and (Right) Highest CT image quality.

Conclusion - Apple phantom: image quality assessment

The DSC, TRE and qualitative results indicate that the image quality does not have a significant impact on the DIR accuracy. It is therefore not justified to use a higher image quality in an attempt to improve the accuracy of the DIR calculations. Increasing the image quality would require an increase in the imaging dose which is not justified. The results also indicate that the DIR calculations have errors for all image qualities and that any DVF calculated would need to be thoroughly evaluated by an expert user before clinical use for contour or

dose propagation, and any contours propagated would also need to be thoroughly evaluated by an expert user. Any contours propagated and used clinically for treatment planning without expert evaluation and editing will very likely be incorrect, this would lead to the possibility of significant errors in dose delivery to the target volume and OAR, and therefore decreasing the tumour control probability and increasing the normal tissue complication which would cause a decrease in the therapeutic ratio.

3.2 Patient Study

3.2.1 General patient image sets: without user input

The 8 patients selected have a range of different treatment sites as shown in table 2.1 and 2.2, this permits a more comprehensive validation of the DIR algorithm because a range of different sites means that the DIR algorithm has to process a range of different images with varying anatomical structures and hence image intensities. The ability of the DIR algorithm to process image data from a range of different treatment sites is implicit in its design, and so it is appropriate that this is assessed.

The AVR was calculated using equation 2.5. The observed range of the AVR in all contours expertly defined on the reference and target CT image sets was from $\simeq 0.5$ to $\simeq 2$, which represents a change in the original anatomical volume of between $\simeq 50\%$ and $\simeq 200\%$. The 22 contours used in this study are a sample of patient contours and give an indication of the general anatomical volume changes you might expect to see clinically.

The DVR was calculated using equation 2.1. The DVR results are listed in table 3.6 and A2. The result of a paired, two-tailed t-test showed that there isn't a statistically significant

difference between the mean DIR-NR DVR and the mean DIR DVR (p = 0.91). The result of a paired, two-tailed t-test showed that there isn't a statistically significant difference between the mean rigid registration DVR and the mean DIR DVR (p = 0.78), indicating that they are essentially equivalent methods of registration. The t-test indicates that using DIR for contour propagation is no more advantageous than using rigid registration, however these results do show that the range decreases when using DIR. The DVR metric only compares the volumes of the expertly defined contour and the propagated contour, it doesn't give any information on the accuracy of the shape and position of the contours which is an important consideration when assessing the DVR results.

	DVR Mean	DVR Range
DIR	1.02	0.54 - 1.83
DIR-NR	1.01	0.53 - 1.85
RR	1.05	0.54 - 2.03

Table 3.6 DVR results of the 22 contours for the general patient image sets: without user input

Looking at the mean DVR values on average the volume of the propagated contours was within 5% of the volume of the expertly defined contours for all registration methods, this indicates good agreement between the propagated contours and the expertly defined contours. However the standard deviation and the range of the DVR was relatively large for all registration methods. The standard deviation was from 0.25 to 0.31 which represents a 25 to 31% difference in contour volume, and the range was from 0.53 to 2.03 which represents 47 to 103% difference in contour volume.

The variability of the DVR indicates that the deformed volumes on the target image set can be significantly different from the expertly defined volumes on the target image set. The difference between the propagated contour volume and the expertly defined contour volume could have significant clinical implications. Any plan that uses a propagated tumour volume,

as opposed to an expertly defined contour, for its optimisation risks having significant under dosing and any OAR volume using a propagated contour risks having significant overdosing. Using propagated contours for planning could potentially result in significant errors in dose, this could have significant implications for treatment, with a potentially lower tumour control probability and higher tissue complication probability, decreasing the therapeutic ratio.

The DSC was calculated from equation 2.2. The DSC results for the DIR-NR, DIR and rigid registration are plotted against the AVR on a scatter plot and displayed in figure 3.20 and tabulated in table A.1. The DSC results for all registration types were plotted against the AVR to evaluate the impact a range of actual anatomical volume changes (i.e. the expertly defined volume changes from the reference to the target) would have on the registration.

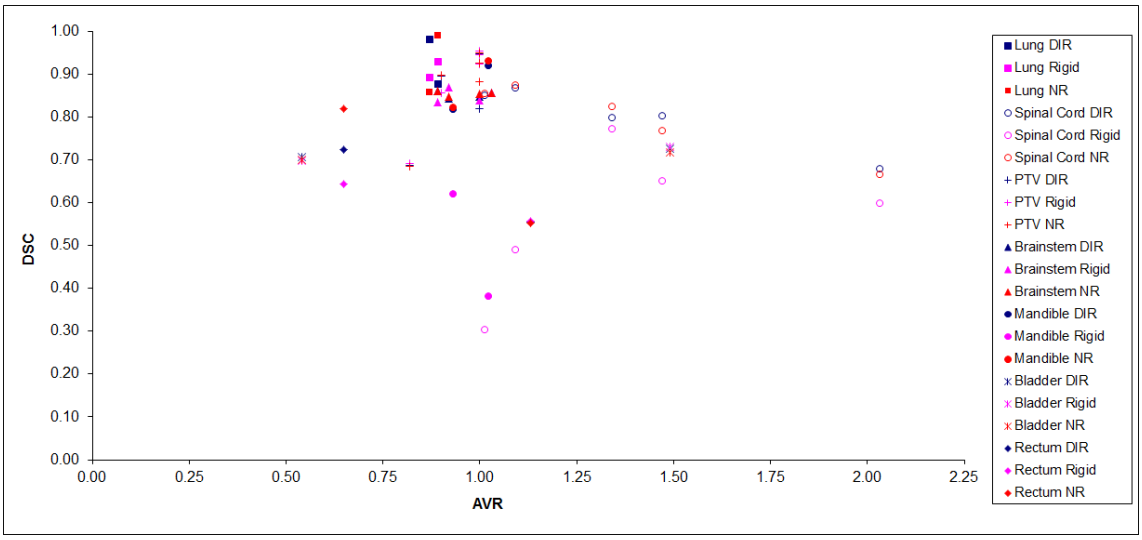


Fig. 3.20 Scatter plot of DSC results against the AVR

From figure 3.20 it is observed that the average DSC for all registration methods trends downward for decreasing and increasing AVR, relative to an AVR of 1. An increase of greater than 1.1 in the relative anatomical volume of the bladder, rectum and spinal cord and a decrease of greater than 0.9 in the volume of the bladder and rectum lead to a decrease in the mean DSC for all registration methods. The mean DSC for all registration methods

is highest for relatively small changes in volume and lowest for relatively large changes in volume as can be seen in figure 3.20. This is expected as a larger difference in volume between the expertly defined contours on the reference image set and the expertly defined contours on the target image set means that the algorithm has to calculate a more complex deformation, as there are more degrees of freedom, which has a greater uncertainty.

From figure 3.20 it is observed that most of the data points cluster between an AVR of approximately 0.9 and 1.1, and a DSC between 0.8 and 1. This is expected as a small change in the AVR should result in a high DSC for rigid registration, assuming there isn't a significant change in shape, and DIR should have a DSC at least as high as the rigid registration DSC if not better, assuming the DIR does not make the registration worse.

The majority of data points which have a AVR of between approximately 0.9 and 1.1 but have a DSC below 0.6 are rigidly registered, this is expected as the anatomical volume has not changed significantly between the reference and the target but the shape of the volume has, and this is not accounted for by the rigid registration. These data points represent the spinal cord and the mandible, which indicates that for these contours the DIR has performed significantly better than the rigid registration. Looking at this region one of the rectum contours has approximately the same relatively low DSC for all registrations, indicating that the shape change of the rectum was too much for the DIR algorithm to bring its DSC up to that of the other propagated contours. The remaining data points that fell outside of the AVR range of 0.9 to 1.1 had a drop in DSC value for all registration methods, however DIR still performed better than rigid registration for the majority of contours.

The mean DSC across all contours for DIR-NR, DIR and rigid registration was 0.82 +/- 0.08, 0.82 +/- 0.07 and 0.72 +/- 0.08 respectively. The mean DSC for DIR-NR and DIR was higher than the mean DSC for rigid registration by 0.1 which is just over one standard deviation for both the DIR and DIR-NR DSC values. The result of a paired, two-tailed

t-test showed that there isn't a statistically significant difference between the mean rigid registration DSC and the mean DIR DSC (p = 0.06). However the result of a paired, two-tailed t-test showed that there is a statistically significant difference between the mean rigid registration DSC and the mean DIR-NR DSC (p = 0.04). It is important to note that the two-tailed t-test was calculated assuming a 95% confidence level, therefore the results are close to the threshold of being considered statistically significant or not. The results indicate that using DIR-NR and DIR is more accurate than using rigid registration alone.

DIR improved the DSC relative to rigid registration alone for the considered contours in 68% of the registrations. In 18% of the registrations the DSC for rigid registration alone was greater than the corresponding DIR DSC. The remaining 14% of DSC values were the same for both DIR and rigid registration. DIR-NR improved the DSC relative to rigid registration alone for the considered contours in 63.7% of the registrations. In 22.7% of the registration the DSC for rigid registration alone was greater than the corresponding DIR-NR of the DSC. The remaining 13.6% of DSC values were the same for both DIR-NR and rigid registration. These results indicate that the algorithm is better at deforming the contours for the majority of cases, however there is a relatively large percentage of contours which are performing worse than the rigid registration, this highlights the need for thorough user evaluation in a clinical setting.

The DIR-NR was assessed because of the known rotational error. The effect of the rotational error depends on the magnitude of the rotational angle, and in this study the rotational angle was relatively small, ranging from 0.53 to 3.74 degrees therefore a large influence on the DIR and hence DIR DSC was not expected [30]. A paired, two-tailed t-test showed that there isn't a statistically significant difference between the mean DIR DSC and DIR-NR DSC (p = 0.83).

The CD was calculated using equation 2.3. The CD was calculated for two registration

methods, DIR-NR and rigid registration. The mean CD (cm) was 0.74 +/- 0.47 and 0.43 +/- 0.58 for rigid registration and DIR-NR contours respectively. The mean CD decreased significantly from the rigid registration to the DIR-NR (p=0.05). These results indicate an improvement in the contour propagation for DIR-NR as a smaller centroid distance indicates a greater agreement with the expertly defined reference contour. However the CD between the DIR-NR propagated contours and the expertly defined contours was smaller than rigid registration for 68% of the considered contours, the remaining 32% of contours had a smaller CD for rigidly propagated contours. Although the results indicate that the DIR-NR performed better than rigid registration for the majority of contours it also indicates that the DIR-NR actually performed worse than rigid registration for almost a third of the contours. These results suggest that it can not be assumed the DIR algorithm gives a better result than the rigid registration and that it can in fact be less accurate, this is an important consideration when being used in a clinical setting as a poorer registration could result in incorrect doses to the target volume and OAR if the propagated contours were accepted for planning.

The centroid distance calculated for DIR-NR is plotted against the DIR-NR DSC for all contours and displayed in figure 3.21. The scatter plot shows that there is a moderate negative correlation, with a correlation coefficient of -0.63, this indicates that there is a tendency for high CD values to go with low DSC values. This is expected, as an increase in the CD indicates poorer registration and a decrease in the DSC indicates poorer registration.

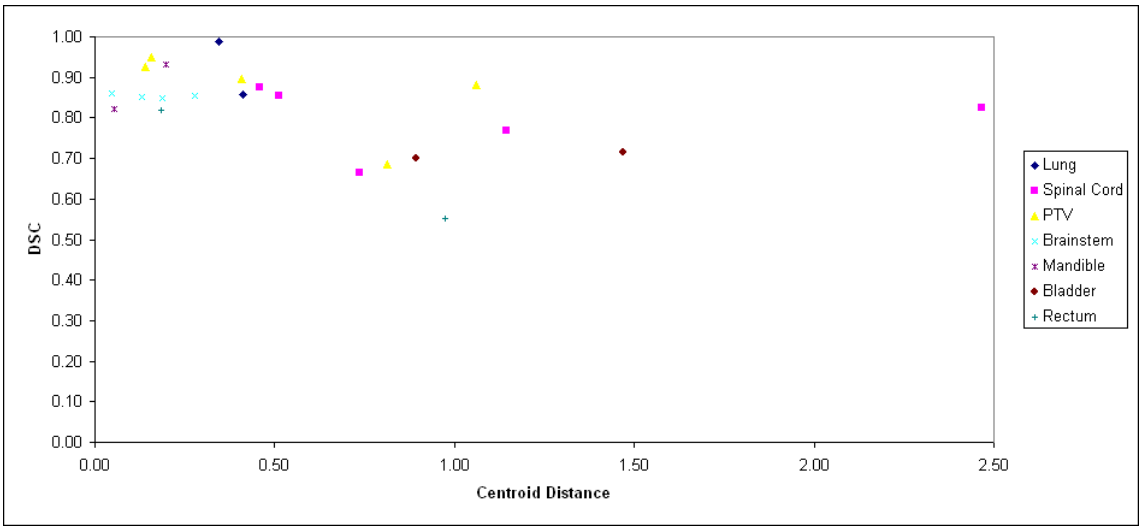


Fig. 3.21 Scatter plot of Centroid Distance results for DIR-NR plotted against the DSC results for DIR-NR

Conclusion - general patient image sets: without user input

The results indicate that there isn't a significant difference between using DIR-NR or DIR, and that DIR-NR and DIR perform better than rigid registration for the majority of cases presented to it. However it's important to note that the results also indicate that for a relatively large number of contours the DIR algorithm performed worse than the rigid registration. This is an important consideration when being implemented in a clinical setting. The contours could not be used for treatment planning without thorough user evaluation and visual inspection by an expert on delineating anatomical structures. If the contours were accepted without thorough user evaluation and used for treatment planning there is a high risk that there would be unintended doses to the target volume and OAR which would impact tumour control and normal tissue complications and therefore the therapeutic ratio.

3.2.2 Head and neck study: without user input

The observed range of the anatomical volume change in the expertly defined contours on the reference CT image data sets to the expertly defined contours on the target CT image data sets, was from 0.5 to 4 times the original volume.

Figure 3.22 is a scatter plot of the DIR DSC results against the AVR, this illustrates the relationship between the actual volume change (as determined from the volumes delineated by an expert) from the reference to the target CT image set and the deformation accuracy. From figure 3.22 it is observed that the DSC values cluster between 0.8 and 1 and the AVR range of approximately 0.9 and 1.1, outside of this region it is observed that the DSC values decrease significantly. All propagated contours have errors associated with them and will therefore require thorough evaluation by an expert user, and possibly editing if they are to be used for clinical use. The results also indicate that any volume change in an anatomical structure from the reference to the target that is larger than approximately 10% will be less accurate than an anatomical change of less than approximately 10% and will therefore be likely to require more editing by an expert user. This indicates that the algorithm is more accurate when it only needs to make relatively small deformations and that the algorithm is less accurate when it is required to make relatively large volume deformations.

The mean DVR for all contours was 0.96 +/- 0.34 with a range of 0.27 to 1.81. The range for the DVR represents a -73% to 81% difference in volume from the expertly defined contour to the propagated contour. Looking at the mean DVR value on average the DIR propagated contours were approximately 4% larger than the expertly defined contours on the target CT image set which indicates excellent agreement. However it's important to also note that the standard deviation and the range is relatively large, indicating that for a number of contours the deformed volume was significantly different from the expertly

defined volume.

The mean DSC across all contours was 0.72 +/- 0.21 and 0.55 +/- 0.23 for DIR and rigid registration respectively. A paired, two-tailed t-test showed that there isn't a statistically significant difference between the DIR and the rigid registration DSC (p=0.29). The DIR DSC has a relatively large standard deviation which is due, in part, to the difficulties the algorithm had at deforming volumes under 20 cm^3 and to the difficulty of calculating accurate deformations for the relatively large volume changes in the contours from the reference CT image data set to the target CT image data set. Figure 3.22 displays the calculated DSC values against the relative anatomical volume change calculated by taking the ratio between the expertly defined contours on the reference image and the expertly defined contours on the target image.

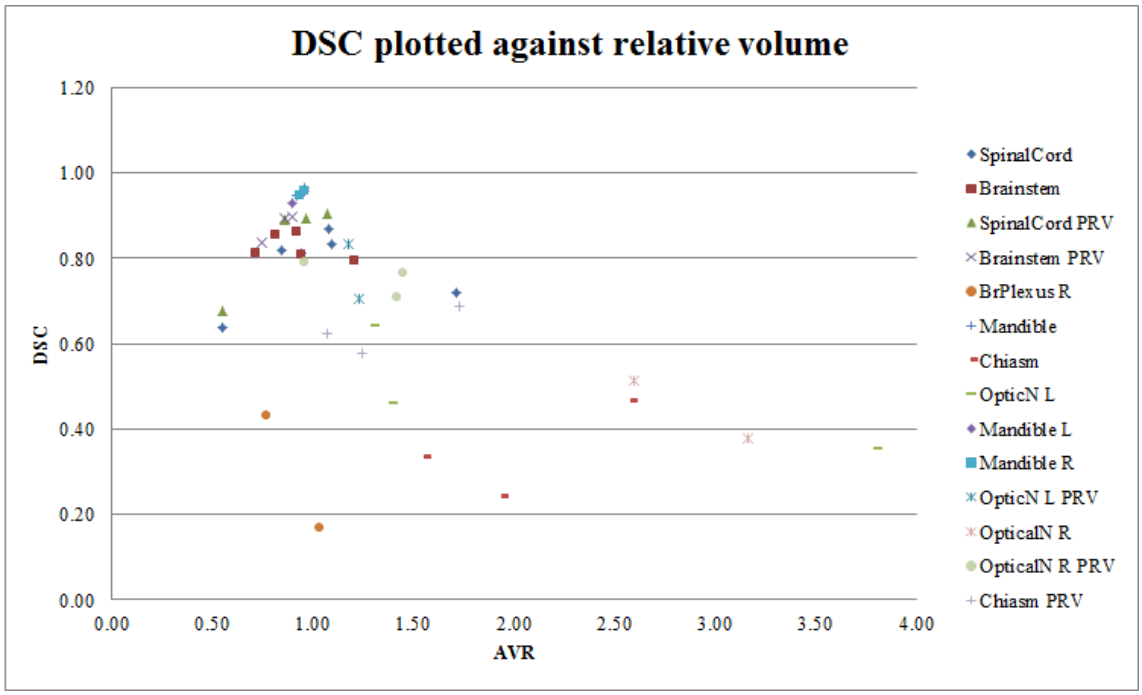


Fig. 3.22 Scatter plot of DSC results against the AVR

The DSC is plotted against the reference contour volume (cm^3) in figure 3.23 this illustrates the relationship between the volume of the contour being deformed and the DSC.

As the volume increases the DSC also increases, this indicates that the deformation is more accurate for larger volumes. For the objects and structures with a volume greater than approximately 20 cm^3 the DIR algorithm had a relatively high DSC indicating an increased deformation accuracy for these contours. This indicates that for smaller volumes the DIR algorithm is relatively poor at calculating deformations.

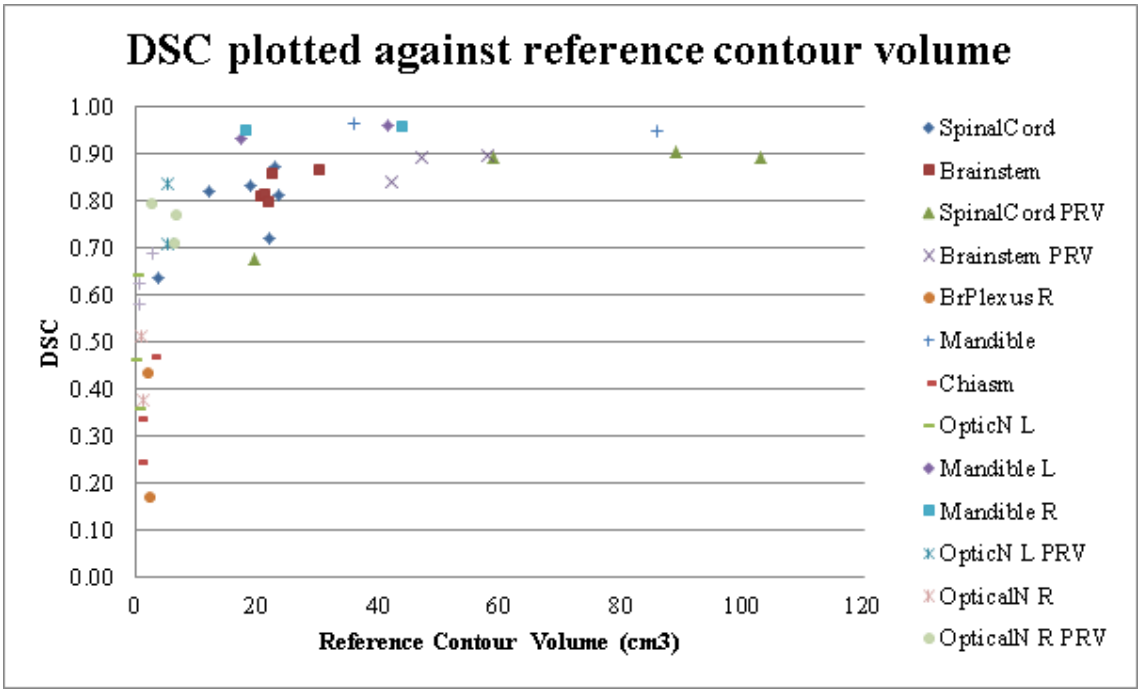


Fig. 3.23 Scatter plot of DSC results against the reference contour volume

The large standard deviation attributed to the DVR and the DSC is due in part to the relatively poor DIR calculations for the smaller volumes. Figure 3.24 shows the comparison between the reference, target and propagated contours for the brainstem on the CT image data set and the colourwash deformation calculated for an individual patient. The DSCs calculated for the brainstem were all relatively high. This is because the expertly defined contour of the brainstem did not change significantly between the reference and the target CT image set, as there wasn't a significant change in the brainstem itself. The DIR algorithm correctly identified that there was no deformation in this region for all patients and hence the reference contour was not deformed significantly yielding a high mean DSC for the

brainstem.

Figure 3.25 compares a single CT slice from the reference and the target CT data set, where the reference, target and propagated contours of the optical nerve are overlaid on both images. Inspecting the images it is clear that the reference right optical nerve is in a slightly different position than in the target. The deformation algorithm did not accurately calculate this shift, as is observed in the images. This may be due to the low contrast in that particular region, it may also be due to the small volume. The DSC is consequently low, matching the qualitative analysis.

The mean DSC for the PTV across all patients for DIR was 0.88 +/- 0.05. A SOP was planned and optimised around the propagated contours to deliver 75 Gy in 35 fractions to the deformed tumour volume for each patient. The dose distribution on the expertly defined contours was then compared and assessed. The mean results for the qualitative dosimetric statistics of the PTV are listed in table 3.7. It was found that the mean $D_{98\%}$ for the expert target volume was -10% of the $D_{98\%}$ planned to the deformed target volume, with a range of -0.34 to -31.69%. Therefore the expert target volume was getting under dosed on average by 10% when the propagated contours were used for clinical planning. This is below the recommended delivery dose of between -5% and +7% according to ICRU Reports 50 and 62 [1][2].

	$D_{98\%}$ (Gy)	D_{mean} (Gy)	$D_{2\%}$ (Gy)	D_{max} (Gy)
Deformed PTV	63.53 +/- 1.84	69.89 +/- 0.23	73.74 +/- 1.16	74.21 +/- 1.28
Expert PTV	57.05 +/- 8.63	69.36 +/- 0.66	73.85 +/- 1.25	74.09 +/- 1.29
Percentage difference	-10%	-0.17 %	0.15 %	-0.17 %

Table 3.7 Mean values for the qualitative dosimetric statistics of the PTV

It was also found that the mean D_{max} for the expert OAR was -3.9% of the D_{max} planned

to the deformed OAR, with a range of -22.84 to 11.56%. Therefore some of the expert OAR volumes were getting a significant overdose with up to 11.56% higher D_{max} than was expected by the plan. This is particularly important for OAR that are serial structures, as a dose that exceeds the tolerance of the structure will cause it to stop functioning. A dose volume histogram (DVH) of one of the patients is shown in figure 3.26. From figure 3.26 it is observed that the dose to the expertly defined PTV is significantly lower than the dose to the deformed PTV, and that the dose to the expertly defined right brachial plexus is significantly higher than the dose to the deformed right brachial plexus, it is also observed that the doses to the other OAR do not match.

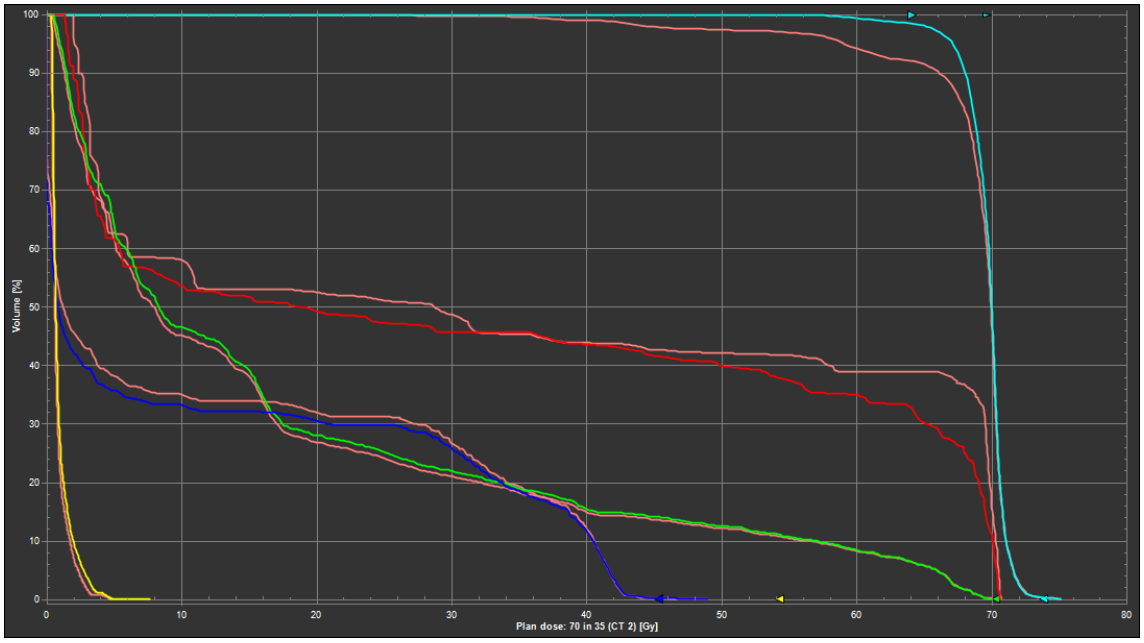


Fig. 3.24 DVH displaying the target volume and OARs for a head and neck patient (Pink is the expertly defined contours, yellow is the deformed brainstem, dark blue is the spinal cord, green is the mandible, red is the right brachial plexus and light blue is the PTV).

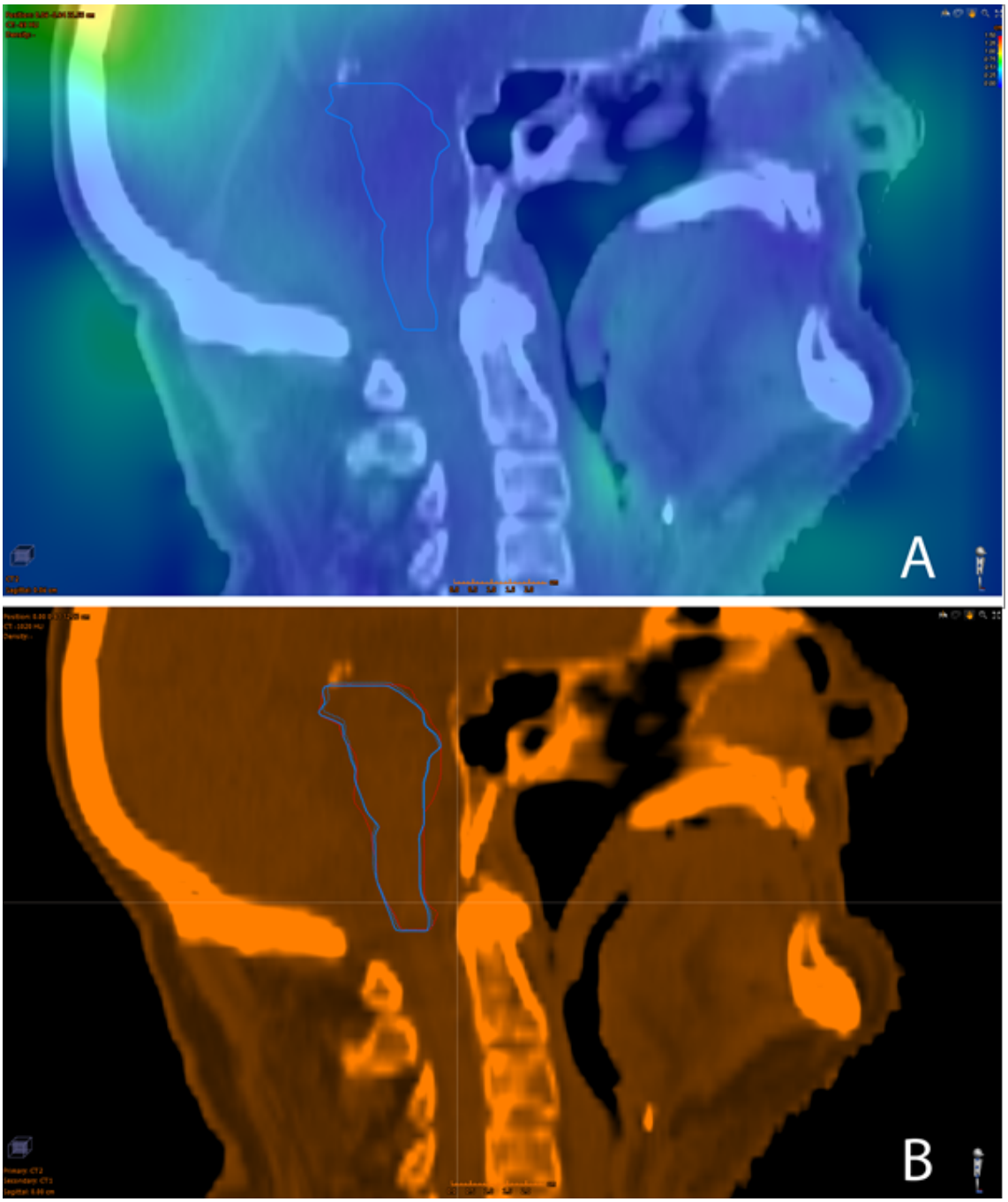


Fig. 3.25 (A) Is a colourwash map illustrating the DVF with the reference contour in blue, the dark blue colour indicates 0 to 0.25mm deformation and (B) is the target image with the expertly defined reference contour (solid blue), expertly defined target contour (solid red) and the propagated contour (dashed blue)

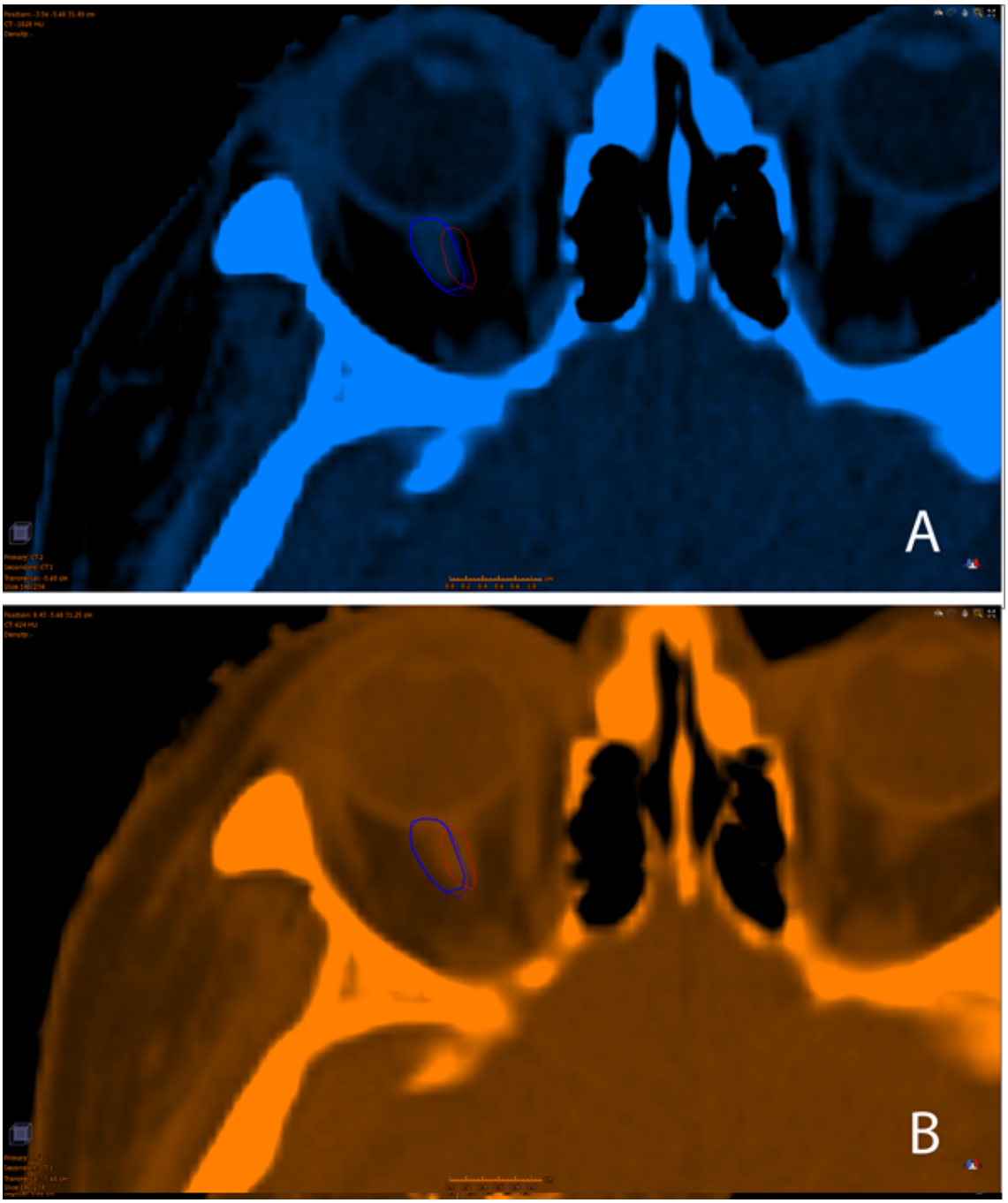


Fig. 3.26 (A) Is A single slice from the reference image and (B) is a single slice from the target image. The expertly defined reference contour (solid blue), expertly defined target contour (solid red) and the propagated contour (dashed blue) are overlaid on both images.

Conclusion - head and neck study: without user input

The results give an indication as to the limitations of the calculated deformation. It was observed that the DIR algorithm is relatively poor at deforming reference contours that have volumes of less than approximately 20cm^3. It was also observed that contours that had an AVR outside of 0.9 - 1.1, and therefore a anatomical volume change of greater than approximately 10% from the reference contour to the target contour, resulted in significantly poorer DIR calculations. Relative volume changes of up to 400% were observed which is really pushing the algorithms ability to deform.

The use of propagated contours for clinical planning and optimisation can potentially cause significant errors in dose compared to the expertly defined contours. The errors may lead to under dosing of the target volume and over dosing of the OAR. The dose errors decrease the TCP and increase the NTCP which in turn lowers the therapeutic ratio and therefore decreases the effectiveness of the treatment, that is curing the malignancy, and increases the chance of side effects. These results show that all propagated contours contain errors which can have significant dosimetric consequences, therefore all propagated contours need to be thoroughly evaluated by an expert user if they are to be accepted for clinical use.

3.2.3 Head and neck study: with user input

The mean DSC for the contoured structures using the DIR algorithm alone was 0.64 +/- 0.24, with a controlling structure it was 0.60 +/- 0.41 and with a focus structure it was 0.54 +/- 0.23. A paired, two tailed t-test for the DSC values of the DIR algorithm alone compared to the DSC values with controlling structures showed that there wasn't a statistically significant

difference between the data sets (p=0.99). A t-test for the DSC values of the DIR algorithm alone compared to the DSC values with focus structures showed there wasn't a statistically significant difference between the data sets (p=0.49).

There were 6 very small structures selected which had a volume of less than 3 cm^3. These 6 structures brought the mean down for the 15 selected structures. The mean DSC for the 6 smaller volumes was 0.38 +/- 0.17, 0.18 +/- 0.24 and 0.33 +/- 0.09 for the DIR alone, DIR using a controlling structure and DIR using a focus structure respectively. A paired, two tailed t-test showed that there wasn't a statistically significant difference between the mean DIR alone DSC and the mean controlling structure DSC (p=0.13) or the mean DIR alone DSC and the focus structure DSC (p=0.54). The DIR accuracy, as indicated by the DSC, did not improve for any of the registration methods when considering the small volumes.

The The 9 larger volumes, which had a volume of at least 12 cm^3, had a mean DSC 0.81 +/- 0.04, 0.94 +/- 0.03 and 0.75 +/- 0.1 for the DIR alone, DIR using a controlling structure and DIR using a focus structure respectively. A paired, two tailed t-test showed that there was a statistically significant difference between the mean DIR alone DSC and the mean controlling structure DSC (p=0.01), but that there wasn't a statistically significant difference between the mean DIR alone DSC and the focus structure DSC (p=0.09). The DIR accuracy, as indicated by the DSC, improved when using a controlling structure, but not a focus structure. Figure 3.27 shows the plot of the initial volume of the contours against the DSC values calculated for the DIR alone, the DIR with a controlling structure and the DIR with a focus structure.

The DSC results indicate that the algorithm performed significantly worse when deforming the small volumes as compared to the large volumes for all registration methods. The small volumes in this study were the chiasm and the right brachial plexus, these are specific contours that are very likely to deform incorrectly, which is important in a clinical situa-

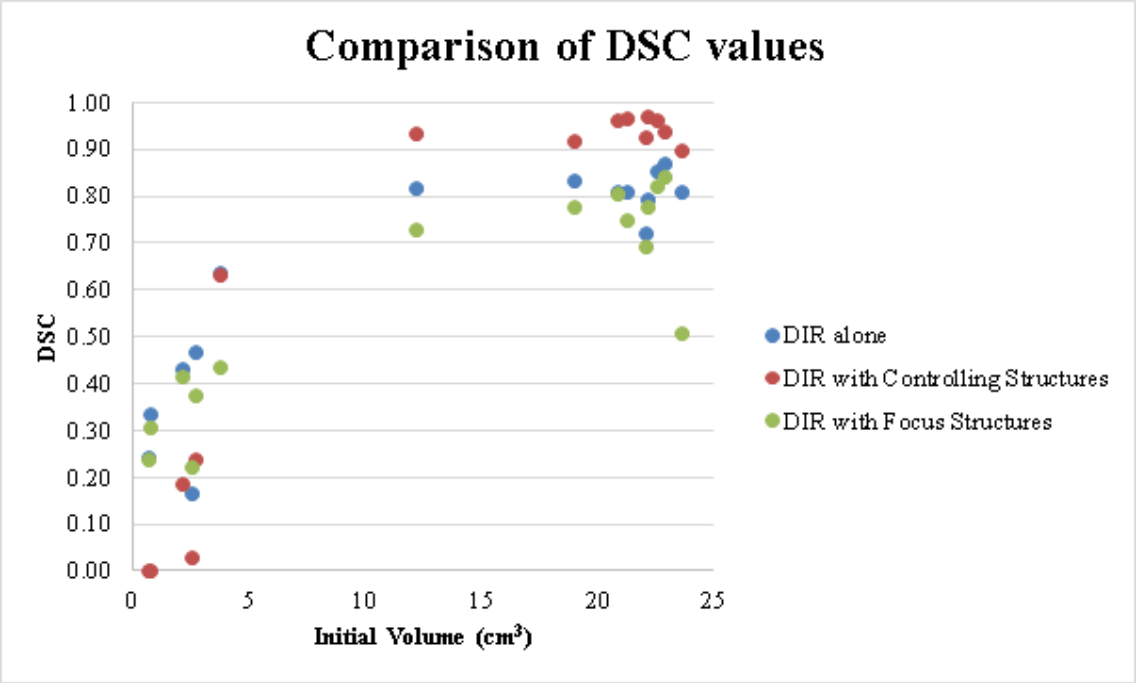


Fig. 3.27 Scatter plot of the DSC values calculated for the DIR algorithm alone, with a controlling structure and with a focus structure

tion. The results indicate that any contours that have a small volume and are propagated by the DIR algorithm need to be thoroughly evaluated by an expert user if they are to be used clinically. This is because they will potentially have significant errors associated with them which will need to be corrected. If the contours are not corrected, there is the possibility that the patient may have the incorrect dose given to the tumour or OAR which will influence the therapeutic ratio negatively.

The results indicate that using controlling structures for larger volumes improves the DIR accuracy, and that using focus structures does not improve the accuracy of the DIR calculations for any volume size. These results indicate that controlling structures should be used clinically for large volumes whenever possible to help guide the algorithm and achieve a higher accuracy in the DVF. However the propagated contours still had errors associated with them, indicating that the DVF was not entirely accurate, therefore all DVFs calculated with the intention of being using clinically must be thoroughly evaluated by an expert user.

The DSC results indicate that the DIR algorithm with a focus structure calculates less accurately than without it for the larger volumes. It was also observed that the calculation time for using a focus structure was considerably faster than without the focus structure, being on the order of seconds as opposed to minutes, part of the reason for the poorer accuracy may be the lower calculation time. The algorithm also has less information in the surrounding volume for its calculation. Figure 3.28 shows the volume that the algorithm selected to calculate its deformation when the spinal cord was selected as a focus structure ROI.

Fig. 3.28 The oblong volume surrounding the spinal cord is automatically selected when the contour is used as a focus structure as is displayed here.

Conclusion - head and neck study: with user input

Using the contour being propagated as a controlling structure enabled the algorithm to more accurately deform the larger contours (> 12cm^3) as compared to DIR alone. The algorithm

was very poor at deforming very small volumes accurately. For small volumes the algorithm performed poorly for all registration methods, that is DIR alone, DIR with controlling structure and DIR with focus structure. The results also indicate that for larger volumes the focus structures are not as accurate as the DIR alone or DIR with a controlling structure.

The DIR algorithm was unable to perform accurate deformation calculations for small volumes even with the added functions. This indicates the possibility of a fundamental limitation in the algorithm's ability to deform contours, that is, a volume beyond which the algorithm breaks down and is unable to accurately propagate contours. This is an important consideration if the DIR algorithm is to be used clinically, as all contours that have a small volume and are propagated will need to be thoroughly evaluated by an expert user. Small volumes are likely to be least accurate when deforming.

3.3 Dose Measurements

3.3.1 Phantom Layers

The dose that the film and ion-chamber received was predicted using RayStation, this was achieved by planning the same beam arrangement used on the machine, and also by planning with the CT image set of the same phantom arrangement used. This gave a calculated planned dose to the film and ion-chamber from which a comparison to the measured film dose, ion-chamber and the deformed dose could be made.

The accuracy of the dose detectors is an important consideration. The dose measurements on the film have an uncertainty of approximately 3%, whereas the ion chamber has an uncertainty of approximately 1.5% [32][3]. The film is significantly less accurate than

the ion chamber, however it provides an approximate value of the dose delivered and has a greater resolution.

Tables 3.8 and 3.9 include the film dose, the plan dose calculated by RayStation, and the deformed dose calculated by the DIR algorithm in RayStation. The dose is measured and calculated at five POIs within the beam and in the plane of the film for comparison. Table 3.8 lists the measured and calculated dose results for the single layer phantom as the target image set. Table 3.9 lists the measured and calculated dose results for the four layer phantom as the target image set.

Point of Interest (POI)	Film measurement (Gy)	RayStation plan calculation (Gy)	Deformed dose (Gy)
1	1.41	1.32	1.28
2	1.43	1.32	1.28
3	1.42	1.32	1.28
4	1.43	1.32	1.28
5	1.41	1.31	1.28
mean	1.42	1.32	1.28

Table 3.8 Seven layer phantom film results

Point of Interest (POI)	Film measurement (Gy)	RayStation plan calculation (Gy)	Deformed dose (Gy)
1	1.36	1.29	1.31
2	1.35	1.29	1.32
3	1.37	1.29	1.31
4	1.35	1.29	1.32
5	1.37	1.28	1.30
mean	1.36	1.29	1.31

Table 3.9 Ten layer phantom film results

The mean film measurement dose for the seven layer phantom was 1.42 Gy, this was 7.6% higher than the planned dose and 10.9% higher than the dose calculated by dose propagation. The mean film dose for the ten layer phantom was 1.36 Gy, this was 5.4% higher

than the planned dose and 3.8% higher than the dose calculated by dose propagation.

Table 3.10 lists the measured and calculated dose results for the ten and seven layer phantom as the target image set.The dose is measured and calculated at the isocentre in the middle of the ion-chamber for comparison.

Phantom	Ion-chamber measurement (Gy)	RayStation plan calculation (Gy)	Deformed dose (Gy)
Ten layer	1	1	1.04
Seven layer	1.03	1.03	1

Table 3.10 Ten and seven layer phantom ion chamber results

The ion-chamber dose measurement for the ten layer phantom was 1 Gy, this matched the dose calculated by the plan on RayStation and was 4% lower than the dose calculated by deforming the dose. The ion-chamber dose measurement for the seven layer phantom was 1.03 Gy, this matched the dose calculated by the plan on RayStation and was 3% lower than the dose calculated by deforming the dose. Therefore the doses predicted by the calculated plans on RayStation were the same as the ion-chamber measurements. The measured doses with the ion-chamber showed excellent agreement between the calculated planned doses on RayStation. This is expected as this is a full buildup region with a relatively small dose gradient, the ion-chamber is therefore more accurate in this region.

The measurements show that the calculated planned doses to the film and ion-chamber using RayStation was significantly more accurate than using the deform dose function. A clinical example of how the dose propagation could be used is where a patient has come in for a re-treatment some months or years after an initial course of radiation therapy. In this case dose propagation could be used to map the dose from the old treatment CT data set to the newly acquired CT data set which would be used to assess the doses that particular structures have already received and hence would influence to what level of dose

those structures could go to in the new treatment. In order to do this the dose propagation needs to be accurate. However these results suggest that it would be more accurate to plan the original beam arrangement on the newly acquired CT data set and then plan the new treatment. dose propagation could also potentially be used clinically for dose accumulation, however these results indicate that the dose inaccuracies are relatively high and so it may not be appropriate to use dose propagation for dose accumulation. The results show that the RayStation calculated planned doses should be used in preference to the doses calculated by the deform dose function and that the deform dose function should be used with caution.

Conclusion - phantom Layers

The film and ion chamber measurements indicated that the calculated planned doses are more accurate than the deformed doses. Therefore the clinical use of dose propagation needs to be approached with extreme caution, as it contains many inaccuracies which could have a significant impact on the treatment of a patient. In order to use dose propagation the user needs to be confident that the DVF calculated by the DIR algorithm is accurate, and would therefore require thorough user evaluation of the DVF, however these results suggest that the errors associated with the DVF are too large for dose propagation to be clinically useful.

Chapter 4

Conclusion

4.1 Conclusion

The quantitative and qualitative results indicate limitations in the DIR algorithm. The results show that the DVF calculated by the DIR algorithm is not entirely accurate and contains errors, which is expected as the calculation of a DVF by a DIR algorithm is considered an ill-posed problem to which there is not a single solution. The errors associated with the DVF translate into errors in the propagation of contours and dose.

It was found that the image quality does not have a significant effect on the DIR calculations. It was also found that contours propagated using DIR are not as accurate as expertly defined contours. The errors in contour propagation were found to be significantly larger for small volumes (<20 cm^3), and for relatively large changes in the AVR. The errors in contour propagation were also found to have a negative impact on the dose distribution to the target volume and the OAR when used as the basis for planning and optimising a clinical treatment. Using propagated contours for clinical planning resulted in an under dose to the

target tumour and an over dose to the surrounding tissue, the dose difference between expertly defined contours and the propagated contours was found to be significant in a number of patients.

From this work it is recommended that all propagated contours are assessed by an expert user to ensure that they are clinically acceptable, if it is determined that they are not clinically acceptable then the contours must be edited before being implemented for a clinical use. Using propagated contours without expert user evaluation and editing resulted in a lower TCP and a higher NTCP which subsequently lowered the therapeutic ratio. Contour propagation has the potential to increase the efficiency of ART by decreasing the re-contouring time required, propagated contours will require expert assessment and editing however this process is likely to take less time than re-contouring from scratch for most patients as the edits are relatively small compared to creating a whole new contour.

The presented studies predominately focused on contour propagation with the use of similarity metrics as a means of assessing the accuracy of the DVFs calculated by the DIR algorithm. dose propagation as a means of DIR assessment was an addition to this. As a result, the contour propagation accuracy is better understood than the dose propagation accuracy. However the results do indicate that the calculated DVFs are inaccurate which has consequences for both contour propagation and dose propagation, in that it causes inaccuracies in both. From this work it is recommended that every DVF is thoroughly assessed if it is intended to be used for dose propagation and that dose propagation should be used with extreme caution.

4.2 Limitations and Future Work

A number of limitations in the methods used to evaluate and assess the DIR algorithm were identified and can be improved upon in future studies. The propagation of ROIs between registered images is good at giving information about the accuracy of the DVF, however evaluating the contours does not give information about the accuracy of the vectors calculated for the voxels inside or outside the contour, only the voxels that are on the edge of the contour. This can give a bias result on the accuracy of the DIR algorithm, as the contours are typically structures that have a relatively high gradient of voxel intensity at the boundary which is easier for the intensity component of the algorithm, the correlation coefficient, to calculate accurate vectors [6]. This can be mitigated by contouring more structures and using structures that do not have relatively high gradients at the boundary. Further study on the effect of the number of contoured structures on the accuracy of DIR should be explored.

The phantom data is useful for simulating a clinical situation that can be related to a real clinical situation. However, a phantom is still a simulation and even the most advanced phantoms can not simulate all the anatomical deformations that a real patient may have. It has been suggested that a possible limitation of using phantom data is that phantoms typically have uniform intensity gradients, as opposed to varied intensity gradients in real patient data, this may produce different results than what might be seen clinically [34][42]. Phantom data is useful however as it gives insights into the accuracy of a DIR algorithm, and for future studies more advanced phantoms with anatomical representations and realistic intensity gradients would be ideal.

A interesting area for future research would be the use of mathematical phantoms, which are image data sets that have a known mathematical DVF applied to them which results in the target image set. The DIR algorithm being assessed is then used to calculate the DVF

between the image data sets, the known DVF is then compared directly to the calculated DVF. This is one of the most fundamental ways to assess the accuracy of the DVF and is an improvement for any future work.

Another interesting area for future evaluation and assessment of the DIR algorithm is requesting expert users to grade contours on a scale of clinical useability by evaluating the contours slice by slice. This would be a useful way to get direct information about the clinical usability of the propagated contours.

The similarity metrics used in this study were the DVR, which is a measure of the relative difference in volume of the expertly defined contour to the propagated contour, and the DSC which provides a number that is directly related to how well the volumes overlap and how similar they are in both size and location. Although the metrics used in this study are good, additional metrics would also be useful. A possible improvement for future studies would be to use a surface distance based metric, this measures the mean and maximum distance from points on the surface of each contour, each point has a corresponding point typically defined by the shortest distance between each contour. The DSC is very good at giving a result which describes the general agreement between two overlapping contours; however using a surface distance based metric would provide more information on where the contours agree and do not agree which is needed as this could potentially be clinically significant.

All the quantitative results that are based on the expertly defined reference contour assume that the reference contour is the gold standard and is therefore the best possible contour for the CT image set. However this is not necessarily true, there is inter-observer variation in the contour delineation and this adds uncertainty to the quantitative results. An improvement would be to get multiple experts to contour the same set of CT images such that the inter-observer variation could be quantified and taken into account when calculating the final deformable results.

References

[1] Icru report 50—prescribing, recording and reporting photon beam therapy.

[2] Icru report 62—prescribing, recording and reporting photon beam therapy.

[3] Agency, I. A. E. (2000). *Absorbed dose determination in external beam radiotherapy: an international code of practice for dosimetry based on standards of absorbed dose to water: Issue 398 of Technical reports series*. International Atomic Energy Agency.

[4] Baskar, R., Lee, K. A., Yeo, R., and Yeoh, K.-W. (2012). Cancer and radiation therapy: current advances and future directions. *International journal of medical sciences*, 9(3):193.

[5] Beyzadeoglu, M., Ozyigit, G., and Ebruli, C. (2010). *Basic Radiation Oncology*. Springer Berlin Heidelberg.

[6] Brock, K. (2013). *Image Processing in Radiation Therapy*. Taylor and Francis.

[7] Bushberg, J., Seibert, J., Leidholdt, E., and Boone, J. (2011). *The Essential Physics of Medical Imaging*. Wolters Kluwer Health.

[8] Byrd, R. H., Nocedal, J., and Schnabel, R. B. (1994). Representations of quasi-newton matrices and their use in limited memory methods. *Mathematical Programming*, 63(1-3):129–156.

[9] Chang, A., Ganz, P., Hayes, D., Kinsella, T., Pass, H., Schiller, J., Stone, R., and Strecher, V. (2007). *Oncology: An Evidence-Based Approach*. Springer.

[10] Collins, D. L., Holmes, C. J., Peters, T. M., and Evans, A. C. (1995). Automatic 3d model based neuroanatomical segmentation. *Human brain mapping*, 3(3):190–208.

[11] Coons, J., Moreman, C., Ragains, M., Scalf, K., Vitaz, T., and Spalding, A. (2010). Improvement of therapeutic index for brain tumors with daily image guidance. *International Journal of Radiation Oncology* Biology* Physics*, 78(3):S277–S278.

[12] Feldmar, J. and Ayache, N. (1996). Rigid, affine and locally affine registration of free-form surfaces. *International journal of computer vision*, 18(2):99–119.

[13] Ferlay, J. and Cancer, I. A. f. R. o. (2001). *GLOBOCAN 2000: cancer incidence, mortality and prevalence worldwide*. IARC press.

[14] Goldman, L. W. (2007). Principles of ct: radiation dose and image quality. *Journal of Nuclear Medicine Technology*, 35(4):213–225.

[15] Goodenough, D. (2004). Catphan 500 and 600 manual.

[16] Halperin, E., Brady, L., Wazer, D., and Perez, C. (2013). *Perez and Brady's Principles and Practice of Radiation Oncology*. Wolters Kluwer Health.

[17] Hill, D. L., Batchelor, P. G., Holden, M., and Hawkes, D. J. (2001). Medical image registration. *Physics in medicine and biology*, 46(3):R1.

[18] Kashani, R., Hub, M., Balter, J. M., Kessler, M. L., Dong, L., Zhang, L., Xing, L., Xie, Y., Hawkes, D., and Schnabel, J. A. (2008). Objective assessment of deformable image registration in radiotherapy: A multi-institution study. *Medical physics*, 35:5944.

[19] Khan, F. (2012). *The Physics of Radiation Therapy*. Wolters Kluwer Health.

[20] Kybic, J., Thévenaz, P., and Unser, M. A. (2010). *Multiresolution spline warping for EPI registration*. International Society for Optics and Photonics.

[21] Latifi, K., Zhang, G., Stawicki, M., van Elmpt, W., Dekker, A., and Forster, K. (2013). Validation of three deformable image registration algorithms for the thorax. *Journal of Applied Clinical Medical Physics*, 14(1).

[22] Lu, W., Chen, M.-L., Olivera, G. H., Ruchala, K. J., and Mackie, T. R. (2004). Fast free-form deformable registration via calculus of variations. *Physics in Medicine and Biology*, 49(14):3067.

[23] Marcu, L., Bezak, E., and Allen, B. (2012). *Biomedical Physics in Radiotherapy for Cancer*. CSIRO Publishing.

[24] Protection, I. C. o. R. (1991). *ICRP Publication 60: 1990 Recommendations of the International Commission on Radiological Protection*. SAGE Publications.

[25] Radiation Oncology, A. C. H. (2013a). Clinical protocol - breast mid isocentre planning. page 3.

[26] Radiation Oncology, A. C. H. (2013b). Clinical protocol - craniospinal with compensator planning. page 4.

[27] Radiation Oncology, A. C. H. (2013c). Clinical protocol - head and neck guidelines for imrt and vmat planning. pages 3–4.

[28] Radiation Oncology, A. C. H. (2013d). Clinical protocol - prostate and prostate bed guidelines for imrt and vmat planning. page 3.

[29] RayStation (2013). Raystation 4.0 user manual. page 180.

[30] RayStation (2014). Field safety notice - medical deivice correction. page 3.

[31] Scarfe, W. C. (2012). A comparison of maxillofacial cbct and medical ct. *Atlas of the Oral and Maxillofacial Surgery Clinics of North America: Digital Technologies in Oral and Maxillofacial Surgery*, 20(1):1.

[32] Shi, C. and Papanikolaou, N. (2006). Analysis of the sources of uncertainty for edr2 film-based imrt quality assurance. *Journal of Applied Clinical Medical Physics*, 7(2).

[33] So, R. and Chung, A. (2010). *Multiresolution spline warping for EPI registration*. IEEE.

[34] Varadhan, R., Karangelis, G., Krishnan, K., and Hui, S. (2013). A framework for deformable image registration validation in radiotherapy clinical applications. *Journal of applied clinical medical physics / American College of Medical Physics*, 14(1):4066–4066. 23318394[pmid] J Appl Clin Med Phys.

[35] Varian (2009). Dose in cbct obi advanced imaging varian.

[36] Wang, H., Garden, A. S., Zhang, L., Wei, X., Ahamad, A., Kuban, D. A., Komaki, R., O'Daniel, J., Zhang, Y., and Mohan, R. (2008). Performance evaluation of automatic anatomy segmentation algorithm on repeat or four-dimensional computed tomography images using deformable image registration method. *International Journal of Radiation Oncology* Biology* Physics*, 72(1):210–219.

[37] Weinberg, R. (2013). *The Biology of Cancer, Second Edition*. Garland Science.

[38] Weiss, E., Wijesooriya, K., Ramakrishnan, V., and Keall, P. J. (2008). Comparison of intensity-modulated radiotherapy planning based on manual and automatically generated contours using deformable image registration in four-dimensional computed tomography of lung cancer patients. *International Journal of Radiation Oncology* Biology* Physics*, 70(2):572–581. e2.

[39] Weistrand, O. and Svensson, S. (2015). The anaconda algorithm for deformable image registration in radiotherapy. *Medical physics*, 42(1):40–53.

[40] Yan, D., Vicini, F., Wong, J., and Martinez, A. (1997). Adaptive radiation therapy. *Physics in medicine and biology*, 42(1):123.

[41] Yeo, U. J., Taylor, M. L., Supple, J. R., Smith, R. L., Dunn, L., Kron, T., and Franich, R. D. (2012). Is it sensible to "deform" dose? 3d experimental validation of dose-warping. *Medical Physics*, 39(8):5065–5072.

[42] Zhong, H., Kim, J., and Chetty, I. J. (2010). Analysis of deformable image registration accuracy using computational modeling. *Medical Physics*, 37(3):970–979.

[43] Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C., Kaus, M. R., Haker, S. J., Wells III, W. M., Jolesz, F. A., and Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index: scientific reports. *Academic radiology*, 11(2):178–189.

Appendix A

Patient Study

A.1 General patient image sets: without user input

Patient	Contour	Rigid registration DSC	DIR DSC	DIR DSC No Rotation
P1	Lung	0.89	0.98	0.86
P2	Lung	0.93	0.88	0.99
P1	Spinal Cord	0.77	0.80	0.83
P3	Spinal Cord	0.49	0.87	0.88
P4	Spinal Cord	0.31	0.85	0.86
P5	Spinal Cord	0.60	0.68	0.67
P6	Spinal Cord	0.65	0.80	0.77
P1	PTV	0.94	0.82	0.88
P3	PTV	0.92	0.93	0.93
P4	PTV	0.95	0.95	0.95
P7	PTV	0.69	0.69	0.69
P8	PTV	0.86	0.89	0.90
P3	Brainstem	0.87	0.84	0.85
P4	Brainstem	0.84	0.85	0.85
P5	Brainstem	0.85	0.86	0.86
P6	Brainstem	0.83	0.86	0.86
P3	Mandible	0.62	0.82	0.82
P6	Mandible	0.38	0.92	0.93
P7	Bladder	0.73	0.73	0.72
P8	Bladder	0.70	0.71	0.70
P7	Rectum	0.56	0.55	0.55
P8	Rectum	0.64	0.72	0.82
Mean	+/- 1 SD	0.72 +/- 0.08	0.82 +/- 0.07	0.82 +/- 0.08

Table A.1 DSC results for general patient image sets: without user input

Patient	Contour	DIR-NR	DIR	Rigid
P1	Lung	1.00	1.00	0.87
P2	Lung	1.00	1.00	0.89
P1	Spinal Cord	0.77	0.77	1.34
P3	Spinal Cord	0.99	0.99	1.09
P4	Spinal Cord	1.07	1.09	1.01
P5	Spinal Cord	0.53	0.54	2.03
P6	Spinal Cord	0.68	0.74	1.47
P1	PTV	1.10	1.09	1.00
P3	PTV	1.07	1.07	1.00
P4	PTV	0.99	0.98	1.00
P7	PTV	1.22	1.22	0.82
P8	PTV	0.99	1.02	0.90
P3	Brainstem	1.17	1.18	0.92
P4	Brainstem	0.98	0.96	1.00
P5	Brainstem	1.00	1.00	1.03
P6	Brainstem	1.16	1.20	0.89
P3	Mandible	1.12	1.13	0.93
P6	Mandible	0.94	0.93	1.02
P7	Bladder	0.69	0.69	1.49
P8	Bladder	1.85	1.83	0.54
P7	Rectum	0.88	0.89	1.13
P8	Rectum	1.11	1.18	0.65
Mean	+/- 1 SD	1.01 +/- 0.26	1.02 +/- 0.25	1.05 +/- 0.31

Table A.2 DVR values for DIR, DIR-NR and rigid registration